Pragmatic Methods for Perennial Ryegrass (*Lolium perenne* L.) Breeding

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Garett C. Heineck

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Eric Watkins and Nancy J. Ehlke
Advisors

Expected September 2019

Acknowledgments

I would like to thank the Department of Horticultural Science and the Department of Agronomy and Plant Genetics for the opportunity to pursue my Ph.D. at the University of Minnesota. The projects presented would not have been possible to complete without the collaboration of many students, faculty, and staff.

I am appreciative of the advice and help received from my advisors Dr. Eric Watkins and Dr. Nancy Ehlke. I also thank Dr. Jacob Jungers, Dr. Brian Steffenson, and Dr. R. Ford Denison for serving on my committee and for the helpful advice received over the years.

For assistance in all of my endeavors in seed production research I am greatly indebted to Mr. Donn Vellekson. His knowledge and dedication to grass seed production systems was critical to this project's success. For technical support on the endophyte related projects I am appreciative of the assistance that was given by Mr. Andrew Hollman. For assistance in structural equation modeling and R coding I am indebted to Kayla R. Altendorf, Dr. Eric Lamb, and Dr. Jeffrey Neyhart.

Finally, I wish to thank my friends and family for their help, understanding, and support throughout my education and during the writing of this dissertation.

Garett C. Heineck

Foreword

This dissertation consists of five chapters, each was written as a stand-alone manuscript. The introduction of each chapter serves, in part, as the introductory literature review. Herein is described a brief summary of each chapter's introduction, methods, and results.

Chapter 1

Crown and stem rust are major diseases of perennial ryegrass (*Lolium perenne* L.). Plant breeders and pathologists often rate rust severity in the field using the modified Cobb scale, but this method is subjective and labor intensive. A novel, open-source system using ImageJ and R was developed to quantify pustule number and area using digital images collected from spaced plants in the field. The computer-processing pipeline included development of training data for prediction of pixel identity using random forest and noise reduction spatial processing. Raters and the computer scored rust severity on plant images of varying complexity including whole-plant (WP), five-leaf (FL), and single-leaf (SL) image series. Computer accuracy was determined using the SL, while the FL series gave insight into the true value of WP severity. Rater ability was assessed using a panel of nine scientists with varying levels of disease rating experience. Results showed rater perceptions of crown rust severity were very consistent across images, but agreement on severity values for a given image were low. Rater consistency for stem rust severity was low and FL scores were not strongly correlated with WP scores (r=0.36, P=0.03) indicating low rater accuracy. The computer-processing pipeline was able to accurately discriminate, count and quantify crown and stem rust pustules on leaf samples. Correlations between computer and the median rater score for crown rust were excellent (r>0.90, P< 0.001) for all image series. Similar to raters, there was lack of correlation between WP and FL series (r=0.20, NS) indicating this technique is limited to leaf or stem samples for stem rust and not applicable to WP. However, the computer-processing pipeline shows promise in replacing visual rating of WP for crown rust.

Chapter 2

Perennial ryegrass (*Lolium perenne* L.) is an important turf and forage species that often becomes infected with *Puccinia coronata* f. sp. *Lolii,* the causal pathogen of crown rust. Disease control through Clavicipitaceous endophytes has been proposed as a potential biocontrol. Two field experiments were designed to determine the influence of native *Epichloë* endophyte infection on natural rust development across a diverse panel of perennial ryegrass germplasm. Experiment 1 used an isogenic population design in which clonal plants infected (E+) or endophyte free (E-) were nested within 14 perennial ryegrass entries. Experiment 2 consisted of isofrequent E+ and E- progeny from isogenic parents. Results showed the endophyte had no consistent marginal effect on crown rust severity across or within entries; however, several isogenic host pairs did show either favorable or antagonistic effects. Despite these sporadic effects, no differences were found between isofrequent family pairs indicating the presence of a host by endophyte interaction. Genotypic and phenotypic data revealed that endophyte isolates were similar within entry indicating that host genotype could be responsible for the highly specific endophyte effect on crown rust. The preponderance of host resistance in mediating rust infection was supported by large broad (> 0.75) and narrow-sense (0.76) heritability estimates for disease severity. These findings support the conclusion that endophyte infection does not play a substantial role in mediating crown rust severity on a population scale.

Chapter 3

Understanding trade-offs between breeding for turfgrass performance and seed production capacity is beneficial for turfgrass breeders working on the improvement of perennial ryegrass. An experiment was designed to identify potential tradeoffs and their mechanistic causes by measuring turfgrass and seed production traits on 20 perennial ryegrass entries grown in two Minnesota environments. Average turfgrass quality scores were not correlated with seed yield at either location. However, several individual turf quality rating dates were moderately correlated with seed yield at both locations ($P < 0.1$). Within these dates, subsidiary turfgrass traits were exhaustively regressed against quality data to identify the optimum combination of variables explaining turfgrass quality. Turf quality was driven by lateral and vertical growth rate, crown rust severity, winter survival, and stemminess. Of these traits it was clear that breeding for increased lateral growth rate and winter survival

would be positively correlated with increased yield. Crown rust severity was nearly perfectly correlated across environments ($r > 0.95$) indicating a favorable response. Vertical growth rate and fertile tiller production; however, were negatively associated with turf quality and were investigated to discover whether breeding for slower growth and steminess would negatively influence seed yield. Although fertile tillers were associated with higher seed yield, there was no relationship with steminess in turfgrass plots. Vertical growth rate was associated with earlier maturity, which is associated with increased seed yield indicating a slight tradeoff. Evidence from the environments and germplasm tested suggest that plant breeders should generally not be concerned with negative tradeoffs between perennial ryegrass turf quality and seed yield.

Chapter 4

Producing adequate seed yield is essential for perennial ryegrass cultivar success in both the turf and forage industries. This study examined the importance of seed yield components across 20 perennial ryegrass entries in both spaced plantings and seed production swards at two locations in Minnesota. Competitive (23 plants m^{-2}) and non-competitive (3 plants m^{-2}) spaced plant nurseries were tested for predictive ability. Structural equation modeling (SEM) was used to determine the indirect and direct influences of yield components on total seed yield. The impetus of this approach was to discover new breeding targets and an ideotype for increasing seed yield. Results showed that when tiller survival was low winterkill was highly influential on seed yield both directly and indirectly through fertile tiller number. Fertile tiller number was more important in spaced plant environments than in swords, but very few differences were found between entries. Spike fertility directly influenced spike yield and was indirectly important for total seed yield. Although the relative importance of individual seed yield components was similar between nursery designs, the competitive design had a superior predictive ability for sward yield via total plant yield and spike fertility.

Chapter 5

Turf-type perennial ryegrass success depends both on adequate turfgrass quality, but also economic seed yield. In most breeding programs, spaced plants are the initial unit of

selection in which observations of related individuals dictate selections of superior germplasm for further testing. As such, spaced plants must be predictive of both seed production and turfgrass growing environments. This research investigated the effectiveness of both standard (3 plants m^{-2}) and competitive (23 plants m^{-2}) spaced plant nurseries as selection environments with respect to sward environments using the same 20 turf-type entries. Seed production, turfgrass and the two spaced plant growing environments were tested at two locations in Minnesota. Turfgrass quality traits were measured in 2017 and 2018 and seed production traits were measured in 2018. Rank correlations for fertility index between the competitive nursery and sward environments at both locations was substantial (r_p = 0.52 and 0.81, $P < 0.05$). Genetic color and crown rust severity were the most prominent variables in models for predicting turfgrass quality. Competition between spaced plants made only a minor improvement of predictive ability for these traits. Overall, increased competition between spaced plants increased the predicative ability (r_s) for both turfgrass and seed production traits. Furthermore, the competitive design takes up less space and often makes measurements and observations much easier for bunch-type grasses.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 1

Using R-based Image Analysis to Quantify Rusts on Perennial Ryegrass

Introduction

Perennial ryegrass (*Lolium perenne* L.) is used globally as a turfgrass and forage. Stem and crown rust, caused by *Puccinia graminis* subsp. *graminicola* and *Puccinia coronata* f. sp. *lolii*, have the ability to drastically reduce the value of perennial ryegrass stands. Crown rust infests leaf tissue and has been shown to reduce dry matter by 37% and green tissue by 94% in susceptible cultivars (Clarke and Eagling, 1994). These losses negatively impact the forage and turfgrass industries by reducing the yield, nutritional quality, and aesthetics of stands (Smit et al., 2005). Stem rust is a major concern to U.S. seed producers who grow 160,000 ha of perennial ryegrass annually (NASS, 2016). Losses in seed yield as a result of stem rust can be as great as 92%, effectively rendering a whole seed crop valueless to the producer (Pfender, 2009). Such losses drive the need for accurate quantification methods to help develop more resistant cultivars through breeding and better management practices.

Rust severity in the field is typically estimated visually as the proportion of plant tissue covered in rust pustules using the modified Cobb scale (Roelfs, 1992). Both diseases have conspicuous macroscopic signs: crown rust pustules are circular and orange while stem rust manifests larger, dark red ovate pustules. The modified Cobb scale has 13 categories on a discrete quantitative scale, each with a pictograph of how the host tissue should look for that category. Breeding programs focused on rust resistance have made measurable gains using the modified Cobb scale (Diaz-Lago et al., 2002). Diagrammatic scales are commonly used in plant breeding programs to increase rater accuracy and precision compared to rating without a scale (Godoy et al., 2006; Nunez et al. 2017; Pedroso et al., 2011). Although signs of these pathogens are distinct, evaluation of rust severity on adult plants is a difficult and time-consuming task. Overestimation of disease severity can be especially problematic when raters are inexperienced. Schwanck et al. (2014) described rater inaccuracy with reference to human propensity to develop systematic bias from personal perceptions. They found that individual perceptions, although inaccurate, were generally applied consistently across experimental units. Visual cues, such as scoring cards, were found to improve consistency. However, absolute agreement between two raters is unlikely, thus it is preferable for a single rater to measure disease severity for an individual trial.

Disease progression is influenced by environmental conditions such as temperature and precipitation as well as the presence and abundance of a virulent pathogen. As a result, disease severity data needs to be collected as a repeated measure, perhaps weekly for several months (Van Der Plank, 1963). This requires a single rater to effectively rate a trial weekly during the time of disease presence, which can be very difficult if breeding populations are large or management trials have many treatments or replicates across several locations.

An alternative approach to repeated visual evaluation in the field is to collect digital images of leaves or plants and utilize a trained image analysis pipeline to quantify and analyze disease severity. Image analysis has already been shown to be an effective method to evaluate disease severity of several foliar pathogens including crown rust of oat using Image Pro Plus software (Diaz-Lago et al 2003) and rice brown spot (*Bipolaris oryzae*) of rice (*Oryza sativa* L.) using Assess software (Schwanck and Del Ponte, 2016). Potential advantages of utilizing digital photography to phenotype plant traits include improved accuracy, increased precision, more efficient use of labor, and the preservation of digital information for future analyses.

Possible concerns with closed-source software include uncertainty surrounding system updates affecting results and the need for annual license renewal. Two open-source analysis tools that can be implemented to rate disease from digital images are ImageJ and R. ImageJ is an image analysis software program with both a graphical user interface and command-line features. The software has useful tools to allow users to select and classify image pixels, which is an important step for developing training data for models. Several image analysis packages have been created for R and some statistical packages offer functional tools and methods to analyze images. One method is to use a machine learning algorithm to build decision trees for individual pixels based on training data. Random forest models accomplish this using a supervised learning method that can be used for classification and regression (Liaw and Wiener, 2002). When a binomial variable is predicted, the proportion of the outcomes is calculated for each pixel in the image. The resulting masked image can be further improved based on the spatial patterns of the feature being extracted. Morphological operations, adaptive thresholding, and segmentation are a few tools that previously have been implemented for plant cell counting (Sankar, et al.,

2014), but may also help determine disease severity on leaf or plant images (Schwanck and Del Ponte, 2016).

The objectives of this study were to: 1) develop three image series and an image analysis pipeline from which the effectiveness of both a human perceptions and computer based rust severity rating could be tested and 2) explore the strengths and weaknesses of using computer-based image scoring compared to visual rating.

Materials and Methods

Approach

Spaced plants in the field were photographed using a digital RGB camera to facilitate observation on whole plants. Subsets of leaves were collected from each plant and photographed in a light box to reduce image complexity and increase resolution. Rater accuracy was assessed by a panel of scientists (referred to as raters). An automated image analysis pipeline (referred to as computer) was developed to assess the potential to replace visual rating. Both the raters and computer scored images of whole plants, five individual leaves from each plant, and an image of a single leaf from each plant. Computer accuracy was validated against both rater scores and manual pixel counting in ImageJ.

Plant material

Images were captured from a diverse collection of 50 perennial ryegrass plants from 14 accessions that varied greatly for rust susceptibility. The 50 selected plants were part of a larger nursery (560 plants) established as spaced plants in a 0.6 m grid in the spring of 2017 at the University of Minnesota Agricultural Experiment Station in St. Paul, MN. Weeds were removed by hand and also controlled with a single application of 0.8 L ha^{-1} 2, 4-dichlorphenoxyacetic acid (Shredder, Winfield Solutions) and 1.5 L ha^{-1} dithiopyr (Dimension 2EC, Dow AgroSciences). Inoculum was spread naturally from rust-susceptible border of the cultivar ‘Linn’ and flanking flowering seed production plots.

Image collection

Three image series (n = 50 each) were developed from the perennial ryegrass spaced plants resulting in 150 total images. Whole-plant (WP) images were taken of whole spaced plants, five-leaf (FL) images were sampled from each WP, and single-leaf (SL) images consisted of a single leaf from each plant (Fig. 1.1-1). Series were designed to cover

a range of complexity, with FL allowing us to make inferences on the WP scale. Images for all three series were taken between the fifth and the eighth of August 2017 during peak rust severity.

Whole-plant images were taken of vegetative plants using a Nikon D300 digital SLR camera mounted on a fixed monopod 80 cm high positioned squarely above the plant. Camera settings were shutter speed of 1/160 s, aperture setting of F8, white balance 5000K, and a focal length of 80 mm. Imaging was conducted between 1200 and 1400 h. Five-leaf and subsequent SL images were taken in a portable light box with the same camera. Camera light box settings were shutter speed of 1/10 s, aperture setting of F8, and a focal length of 80 mm. Images were taken over the course of 2 days by mounting five leaves on a matte black board using badge clips and Elmer's tack and stick adhesive strips (Elmer's Co, Atlanta, GA) (Fig. 1.1-1B). Leaf sampling was conducted in a cross pattern from each plant to limit bias and care was taken to avoid dead leaves (Fig. 1.1-1A). Leaf samples were imaged in the field immediately after placement on the imaging board. The imaging box was lit by 150 cm of high color rendering index LED light strip with an output of 221 lumens per 0.09 m^2 and a 4000K color value (superbrighLED.com). The lights were powered by a rechargeable Talentcell 12-volt lithium ion battery pack (talentcell.com). Single-leaf images were cropped randomly from the FL images (Fig 1.1-1C). Images were saved in JPEG format with an image size of 4288 x 2848 pixels.

Visual scoring of images

Nine scientists representing several plant science labs at the University of Minnesota rated disease severity on the image series indoors. Of the nine raters, six were either trained plant pathologists who focus on *Puccinia* spp. or scientists who routinely scored rust as part of a research objective or project. The remaining three had no experience rating rust but had routinely applied visual scales to plants in the field. A review of the literature shows that rater panels vary greatly in size, from three to twelve (Karcher and Richardson, 2003; Godoy et al., 2006; Schwanck and Del Ponte, 2016; Nunez et al. 2017; Pedroso et al., 2011), thus we determined nine would give an adequate representation of what could be commonly found within the plant sciences for rating rusts on a visual scale.

Before rating images, three calibration images were used to train raters *a priori* on rust rating using the modified Cobb scale. Raters were allowed to use the diagrammatic

scoring cards described by Roelfs (1992) during the rating process. The rating was conducted as a fully crossed design where each rater scored each image once within each series after being trained on the calibration images. Raters were not permitted to discuss scores during the rating time period and recorded their own scores.

Inter-rater reliability was assessed using intraclass correlation (ICC) estimation. Because of the fully crossed design, a two-way model was used and rater by image interaction was incorporated into the model. Inter-rater reliability was determined for consistency of rankings. The R package 'irr' was used to measure ICC and obtain standard errors around the estimates (Gamer et al., 2012). Reliability will be discussed according to Cicchetti (1994) who determined that ICC values less and 0.40 to be poor, 0.40 - 0.59 fair, 0.60 - 0.74 good, and 0.75 - 1.0 excellent.

Image analysis pipeline

The image analysis pipeline (Fig. 1.1) was developed using open-source programs and included four steps: 1) development of training data, 2) application of training data to fit a random forest model, 3) creating masked images using the random forest model, and 4) processing the mask using several morphological and transformational techniques. ImageJ was used for its graphical user interface, which was useful in generating training data (Schindelin et al., 2012). R was used for the rest of the image analysis pipeline (Version 3.5.0; R Core Team, 2018). Key packages within R included 'randomForest' and 'EBImage' (Liaw and Wiener, 2002; Pau et al., 2010).

Developing training data

Training data were developed from a subset of four images from each series. The data consisted of groupings of individual pixel RGB values typifying categories within the image such as rust pustule or healthy leaf tissue (Fig. 1.1-1D). Red, green, and blue values were determined using ImageJ function 'RGB measure', which extracted and saved values as comma separated value files (Fig. 1.1-2). Initially, all data generated in ImageJ were saved into a training palette directory. From this directory "mixes," or combinations of RGB values used to detect and separate image features, were combined into a "training palette." For example, a palette to detect stem rust may include more stem rust RGB mixes than a palette for crown rust detection. In this case, overloading the model with crown rust-

related mixes may reduce accuracy. This is to say not all training mixes were needed for every training palette but were added as needed to optimize the model.

Fitting the random forest model

Random forest models are collections of decision trees grown from the training data to classify a single feature (Fig. 1.1-2). The number of trees and nodes from each branch can be selected to influence explained variance and processing time. We found that growing 100 trees (*ntree*) and sampling one explanatory variable per node (*mtry*) on each tree reduced error. An analysis was conducted to quantify the optimal number of pixel measurements to explain the greatest amount of variation (R^2) in a training palette designed to separate foreground from background (Fig. 1.1-3). Training data size was optimized by collecting 1200 biological (foreground) and 1200 non-biological (background) pixels across four representative images within each series. Random forest models were fit for a range of random sample sizes, n=50 to n=1200 and then bootstrapped for 100 iterations. The resulting median R^2 value from each model was regressed against sample size using a square root quadratic model. Two metrics determined optimal sample size: 1) sample size to attain an R^2 of 0.90, and 2) the sample size that maximized R^2.

Image processing using randomForest and EBImage

Images from each series were processed using trained random forest models. Each image was first cropped in R to both standardize pixel number and reduce processing time. The image arrays were then vectorized and bound to a data frame containing the x and y coordinates of each pixel from the original matrices. The random forest model was applied to the data frame where each row (pixel) was assigned a probability. A new random forest model was fitted for each feature of interest (foreground, crown rust, and stem rust). Global thresholding was set at a probability of 0.80 throughout all analyses to reduce the number of erroneous classifications (Fig. 1.1-3).

Further image processing was then conducted in the R package 'EBImage' within the same script. A dilation of the image using a small kernel size followed by adaptive thresholding of the images filled gaps that the random forest model missed (Fig. 1.1-3C). The final product of this processing pipeline gives the user an image free of most background (soil, rocks, etc.) pixels, while retaining the foreground (plant and rust). Isolating pustules required both adaptive thresholding trained on disc shapes and a

watershed transformation to separate objects in close proximity (Fig. 1.1-4). EBImage functions were used to count and quantify the area and shape of pustules.

Image analysis validation

Four tests were conducted to validate the computer-processing pipeline: 1) prediction of severity values on the published modified Cobb scale, 2) estimation of foreground pixel area, 3) ability to quantifying stem and crown rust pustule number on SL, and 4) measuring crown and stem rust severity estimates on SL.

The computer was trained to predict rust severity on the same Cobb scoring cards used by the raters. Cobb scales were scanned as high-quality JPEGs using 48-bit color with 800 DPI resolution and opened in ImageJ. Both image rating styles (small and large pustules) were scanned for analysis. Training data were generated as previously described and applied to a random forest model. The ratio of foreground (colored dots in this case) to total pixels was correlated with the actual percentage of "tissue" occupied by rust uredinia on the published scale.

Each image contained a foreground and a background, which needed to be separated to improve processing time and make accurate calculations of rust severity (Fig. 1.1-3). Pixel areas were validated by taking manual measurements of each image. Whole-plant images were calculated using ImageJ and HSB thresholds. When each image was appropriately thresholded, the number of pixels remaining was recorded. For FL and SL, polygons were drawn around each leaf and the number of pixels within each polygon was recorded. Simple linear regression determined the correlation coefficient between computer and manually collected values.

The computer's ability to accurately quantify crown and stem incidence rust *per se* was evaluated by training models for both diseases to quantify pustule number. Adaptive thresholding and a watershed transformation were used to increase pustule count accuracy on SL images (Figure 1.1-4). The count data generated from this process was correlated with the average value from manual counts collected by two scientists. Data were square root transformed to meet the assumptions of simple linear regression.

Image analysis of total rust severity was conducted in parallel with the pustule count on SL images. Crown and stem rust severities were summed for this test and compared to manual measurements. Total pustule area was manually quantified using ImageJ to draw

polygons around all pustules on each of the SL images. Computer severity was then regressed against the average value from two manual calculations. Data were square root transformed to meet the assumptions of simple linear regression.

Comparing rater and computer scores

Using optimized training data, the computer scored all images for both crown and stem rust severity. The median rater score for each image was compared with computer scores. Rater and computer scores for each series were correlated independently and against each other; data were square root transformed when necessary to meet statistical assumptions. Series compared within scoring methods (rater or computer) gave insight into the accuracy on a WP scale, meaning that correlation strength between FL and WP series and for stem and crown rust indicated accuracy on the WP scale. Series compared between the computer and rater gave insight into the agreement between the two methods.

A validation test was also conducted between visual estimates taken in the field and computer predictions for crown rust. Field ratings were conducted by a single rater in the summer of 2017 on all 560 plants in the nursery. Images were also taken of each plant at the time of visual rating. Severity measurements were compared for a single rating and ratings over time using area under the disease progress curve (AUDPC) method (Roelfs, 1992).

Results

Assessment of rater variability and accuracy

Intra-class correlation values approaching 1 indicate optimal inter-rater reliability, or severity score consistency based on ranking across the images within series. Crown rust scores across all image series had excellent ICC coefficients above 0.85, with only small increases as image complexity decreased. Stem rust scores for WP showed fair agreement while FL and SL showed good agreement (Fig. 1.2A).

Single-leaf images were the least complex; however, rater scores showed high levels of variability. As severity approached 50% on the modified Cobb scale, the distribution of the rater scores was broadest (Fig. 1.2B). However, variability dramatically decreased near the maximum (crown rust only) and minimum extremes of the scale (0 and 100% severity). There were low, but still visually detectable, levels of stem rust compared

to crown rust. The median coefficient of variation (CV) of the 50 images within each series for both crown and stem rust was greater than 50%, indicating large amounts of variability for both diseases. As image complexity increased (e.g. more leaves were observed at once) so did median CV for stem rust (Fig. 1.2C).

Whole-plant severity was estimated by the mean severity across the five leaves sampled from each plant. Correlation coefficients were high for both crown and stem rust between rater scores on FL and SL. The relationship between FL and WP was variable: crown rust scores showed a strong relationship whereas stem rust had very weak correlation (Table 1.1).

Optimizing training data for random forest

The proportion of variance in the training data explained by the model after bootstrapping was described by the correlation of determination (R^2). Whole-plant models reached an R^2 of 0.90 with 500 samples while FL and SL required 250 (Fig. 1.3). The maximal sampling points for WP, FL, and SL were not statistically different and were maximized at about 875 data points. It is important to note that sampling to the maximal point only increased R^2 by about 5% as all models plateaued near 0.95.

Validating the analysis pipeline

The image analysis pipeline was able to predict the modified Cobb scale with only small deviations in the exact area covered by rust pustules. This slight difference may have been due to variability introduced during scanning or slight errors in the scale itself. The correlation coefficient between the Cobb and computer severity values for both small and large pustules were nearly perfect ($r = 0.99$, $P < 0.001$).

Pixels of biological material in all series was separated from the background of each image and then compared against manually measured values. This validation step isolated the entire plant from the soil in WP images and leaves from the imaging plate in FL and SL. Correlation coefficients for estimating the proportion of biologically related pixels in all three image series were 0.97, 0.99, and 0.98 for WP, FL, and SL respectively ($P < 0.001$).

The analysis pipeline could quantify and distinguish crown and stem rust pustule numbers with a high level of accuracy (Fig. 1.4); for example, the pipeline capably predicted crown rust pustule number ($r = 0.93$, $P < 0.001$). The slope of the regression line

for crown rust pustule number was nearly 1 (β = 1.01), meaning that the predicted count was very similar to that of the manual count. The computer did an adequate job predicting stem rust pustule number ($r = 0.87$, $P < 0.001$); with a slope also close to 1 (β = 0.91).

There was a high correlation between predicted total area and manually calculated infested area ($r = 0.96$, $P < 0.001$); the slope of the model was less than 1 (β = 0.80), which indicated that the estimates were larger than actual infestation (Fig. 1.4). Median total rater proportions were very similar to the computer prediction ($r = 0.94$, $P < 0.001$). The correlation between computer FL and WP scores for crown rust was fair (0.84, P < 0.001). However, the correlation between stem rust severity scores were not significant (Table 1.1).

Correlation and consistency between median rater and computer severity scores

Although there were high levels of rater variability for crown rust scores (CV = ~50%), the ICC values strongly indicated excellent consistency and therefore the median rater score was used for analysis. Computer severity ratings were calculated based on the percentage of tissue infected with uredinia. These values were regressed against rater scores that were corrected to account for actual tissue infested with rust (100% severity = 37% actual tissue infested). Computer scores for all three image series were highly correlated with rater scores for crown rust (Table 1.2). Slope estimates for FL and SL were both nearly 1:1 in nature (β = 0.91). The regression for WP had a slope of 1.25, meaning that rater scores generally were greater than that of the computer (Supplemental Fig. 1.1).

Overall, correlations between computer and rater scores for stem rust were lower than for crown rust. Single-leaf scores between computer and the median rater score were superb for crown rust (0.89, P < 0.001) and fair for stem rust (0.79, P < 0.001). As image series increased in complexity from SL to WP, correlations decreased with no significant correlation found for WP (Table 1.2). Interactions were explored on an individual rater basis due to lower levels of rater reliability for all image series (Fig. 1.2A). Although correlations between the computer and individual raters differed, there was no one rater who rated better or worse than the rest for FL and SL (Table 1.3). Furthermore, there also seemed to be no obvious trend between rater experience level and strength of correlation for any series.

Discussion

Diagrammatic rating scales have been used to measure rust severity since the mid 1900s. The modified Cobb scale, in particular, has been used in breeding programs as well as agronomic testing for efficacy of fungicides (Welty and Barker, 1993; Wanyera et al., 2009). More recently, image analysis has been tested as a substitute for visual rating. Typical digital images encompass three wavelengths (RGB), but also can be extended to hundreds of bands using hyperspectral imaging. A difficult question that often hampers the use of image analysis on a large scale is: "How comparable is computer rating to visual scoring?" Intuitively it would seem that a computer should be more capable of accurate and precise rating, however proving this is a difficult task. The three image series used in this experiment investigated this question for crown and stem rust by: 1) calculating consistency and agreement of raters through ICC and variance measures; 2) assessing the accuracy of rater and the computer scores by comparing FL and WP values; and 3) exploring the relationship between rater and computer scores.

When observing vegetative grass plants, a single lamina is the smallest and most conspicuous unit of examination and a sufficiently large sample estimates the mean severity of the whole plant for leaf diseases. Considering visual ratings in particular, a single leaf is the observational scale that allowed raters to best estimate the objective truth. However, heterogeneity of lesion distribution across the leaf surface likely influences rater accuracy (Bade and Carmona, 2011). Images are complex due to both the number of possible RGB profiles and the spatial arrangement of profiles; these features may have provoked raters to differentially interpret signs of disease. Rating even single leaves for both crown rust and stem rust proved challenging; for instance, as severity approached 50% rater scores became highly variable (Fig. 1.2B and C). These findings imply that rating severity as a repeated measure should be done by as few raters as possible to reduce experimental error. Although agreement on precise values was low, consistency among raters was generally high, especially for crown rust (Fig. 1.2A). High consistency among rater scores indicated that the general perception of a heavily infected plant will remain consistent, but the exact value associated with infestation will likely change based on rater.

Proof of computer accuracy was determined based off of the following results: 1) the analysis pipeline itself was able to accurately assess the modified Cobb scale, 2)

computer-predicted tissue proportions were nearly identical to manually calculated values (the denominator for severity calculation), 3) correlations between the manual and computer measures for pustule counts were substantial ($r > 0.75$, $P < 0.001$), and 4) computer estimates for total rust severity explained 92% of the variance for manual predictions (Fig. 1.4). Finally, correlations between computer and median rater severity scores were always strong for SL images (Table 1.2). These findings are in contrast to Bade and Carmona, (2011) who used Assess 2.0 software to count pustules and quantify severity on maize leaf samples infected with *Puccinia sorghi.* They found a low amount of variation explained between hand and computer measurements for both pustule number ($R^2 = 0.49$) and severity ($R^2 = 0.48$).

The analysis method described herein was based completely on open-source programs and can be implemented with minimal experience coding in the R environment. Although not as rapid to analyze a single image or as user friendly as ImageJ, R allows the use of a random forest algorithm ('randomForest'*)* and unique processing functionality ('EBImage') with immediate downstream data analysis potential (Pennekamp and Schtickzelle, 2013) which can be used to easily analyze a large number of images. As a comparison with R, we used the built in ImageJ manual threshold functionality to quantify total rust in relation to manual measurements on SL images. Results show that ImageJ correlations were similar to R ($r = 0.94$ vs. 0.96 respectively). Although the correlation coefficients were nearly identical, ImageJ overestimated severity on SL more so than R ($\beta = 0.63$ vs. 0.81) and took far longer to accomplish on a large set of images (Supplemental Fig. 1.2).

Appropriate training data is essential for our system to work and time must be taken to collect additional data to make accurate models for new image series. We found that WP images required larger training datasets compared to SL and FL images and sufficient sampling achieved R^2 values near 0.95. We found that there were observable decreases in efficiency with higher amounts of sampling of the same profile. For example, increased sampling on the same plant had little effect on R^2 (Fig. 1.3). However, it is important to note that a large R^2 does not necessarily mean that the model will separate background from foreground in practice and adding additional profiles of interest will likely alter the training palette accuracy. For example, creating a new mix of crown rust pixels for different plants

and adding them to a palette that already has an optimized R^2 may increase accuracy on a larger set of images.

These results clearly show that our automated analysis pipeline can replace human rating of vegetative grass plants for crown rust severity rating. Utilizing the median rater score allowed us to gain insight into general rater perception of rust on a standard rating scale. Computer severity scores were strongly correlated with median rater scores across image series for crown rust ($r > 0.90$, $P < 0.001$). Computer crown rust severity scores for WP and FL were also highly correlated. This evidence, combined with the computer accuracy validation, supports the supposition that this system can replace human rating in the field. Computer scores for WP were generally lower in severity than that of the median rater scores ($\beta = 1.25$), meaning that computer scores for spaced plants may tend to be more conservative than the typical rater. This, of course, could possibly be corrected for by adjusting the global thresholding parameters in the script. Five-leaf and SL scores were very nearly 1:1, meaning the proportion of tissue infested with uredinia were nearly the same.

Stem rust proved much more difficult to quantify for both the rater panel and the computer. Although disease severity could be quantified by the computer on leaf tissue (Fig. 4B), correlations between WP and FL were insignificant for computer scores and low for rater scores ($r = 0.36$, $P = 0.03$). Furthermore, computer and rater scores for WP were not correlated. This implies that raters and the computer were both inaccurately quantifying stem rust, but in different ways regardless of rater experience (Table 1.3). It is possible that stem rust is more difficult to rate on WP because of some leaves shading necrotic leaf tissue, which created an extensive range of RGB values similar to that of stem rust pustules. Based on the variation explained for pustule counts and strong correlations between SL and FL severities FL samples were relatively unaffected by this issue, at least for computer scoring (Fig. 4B and Table 1.1). This may have been due to FL images being taken under consistent high light conditions. The utility of this analysis pipeline for stem rust should not be completely dismissed due to economic importance in seed production systems. Measuring stem rust on stem samples collected from the field would be an effective way to utilize this system, especially if there were few experimental units in the study.

Efficiency of this system on a large scale for either WP or FL images compared to visual rating in the field is an important consideration for both breeders and agronomists. Speed of visual rust rating on WP was recorded in the nursery that was used in this study (n = 560). Visual scoring took about 6 seconds per plant compared to imaging the WP which only took about 4 seconds per single plant. However, obtaining FL images which required harvesting leaf samples, positioning them, and capturing the image, took 140 seconds per plant. Schwanck and Del Ponte, (2016) noted that leaf sampling and lesion counting is labor intensive and that visual estimates of severity may be faster. This was certainly the case in our study, however, gains in precision and accuracy for diseases such as stem rust at the cost of up-front time expenses might be worthwhile for many researchers.

Visual scores taken on the 560 plants in the field were used as an additional validation for crown rust. The subset of 50 images scored by the panel was highly correlated with field ratings on the same 50 plants ($r = 0.89$, $P < 0.001$). Visual field ratings from the whole data set were compared against computer scores for crown rust and found a similar correlation coefficient ($r = 0.87$, $P < 0.001$). These findings are similar to Bai et al. (2018) who found a strong correlation between field and indoor rating measures (r = 0.65 to 0.85, $P < 0.001$), as well as high accuracy between indoor measures and computer scores (56.3 to 93.1%) for iron deficiency of soybean (*Glycine max* L.). We found the rater panel and computer scores to be more highly correlated with each other ($r = 0.94$, $P<0.001$) than with field scores taken by a single individual, but in both cases, we found the computer generally had lower scores than the visual rater (Supplemental Fig. 1 and 3).

Conclusion

An open-source, automated image analysis system was developed and accurately predicted crown rust across three levels of image complexity, with spaced WP being the most complex and important from a breeding and management standpoint. It is unlikely that this system in its current state can accurately quantify stem rust on a WP scale. However, leaves or stems collected from infested plants can be used to both quantify severity and count pustules.

Other open-source R packages, specifically the caret package (Kuhn, 2008), includes classification algorithms such as support vector machines and k-nearest neighbor functions that may offer similar or improved predictions depending on the application (Naik et al., 2017). Also, future research should implement a light box for WP imaging which may increase the ability to detect stem rust by increasing light intensity and reducing shadows.

Tables

Table 1.1 - Correlations between all three-image series: WP = whole-plant, FL = five-leaf sample, SL = single-leaf sample. Each experimental unit (plant) is represented by each series so that SL was a sample of FL and FL were a sample of a WP. Rater correlations were conducted with the median panel score for each image.

	Rater crown rust (*r*)			Computer crown rust (*r*)		
	WP	FL	SL	WP	FL	SL
WP	-	***	***	-	***	***
FL	0.89	-	***	0.84	-	***
SL	0.86	0.93	-	0.82	0.94	-
	Rater stem rust (*r*)			Computer stem rust (*r*)		
WP	-	*	***	-	NS	NS
FL	0.36	-	***	0.20	-	***
SL	0.51	0.70	-	0.09	0.85	-

*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels.

Table 1.2 - Correlations between computer and median rater score for each image series: WP = whole-plant, FL = five-leaf sample, SL = single-leaf sample.

	Computer severity score	
Rater severity score	Crown rust	Stem rust
WP	0.94 ***	NS
FL	0.94 ***	0.65 ***
SL	0.93 ***	0.79 ***

*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels.

Table 1.3 - Correlations between individual raters and computer score for stem rust. Rater experience is also provided to show any potential influence that may have had on correlation coefficients. Correlations between all three-image series are shown: WP = whole-plant, FL = five-leaf sample, SL = single-leaf sample.

Stem Rust	Rater experience level †	Computer score		
		WP	FL	SL
rater 1	2	NS	0.67 ***	0.83 ***
rater 2	2	0.26 .	0.71 ***	0.80 ***
rater 3	1	NS	0.64 ***	0.70 ***
rater 4	2	NS	0.63 ***	0.72 ***
rater 5	2	0.31 *	0.64 ***	0.82 ***
rater 6	1	NS	0.72 ***	0.79 ***
rater 7	3	0.41 *	0.57 ***	0.76 ***
rater 8	3	NS	0.70 ***	0.79 ***
rater 9	1	NS	0.59 ***	0.69 ***

*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels.

† Rater experience level was self-assigned before rating began: 1 = to the rater having very little experience rating rust, but experience applying qualitative or ordinal scales to plant tissue; 2 = to the rater scoring rust severity routinely as part of a research objective or project; and 3 = to the rater having studied rusts and has rated often as part of a project main objective or job.

Figures

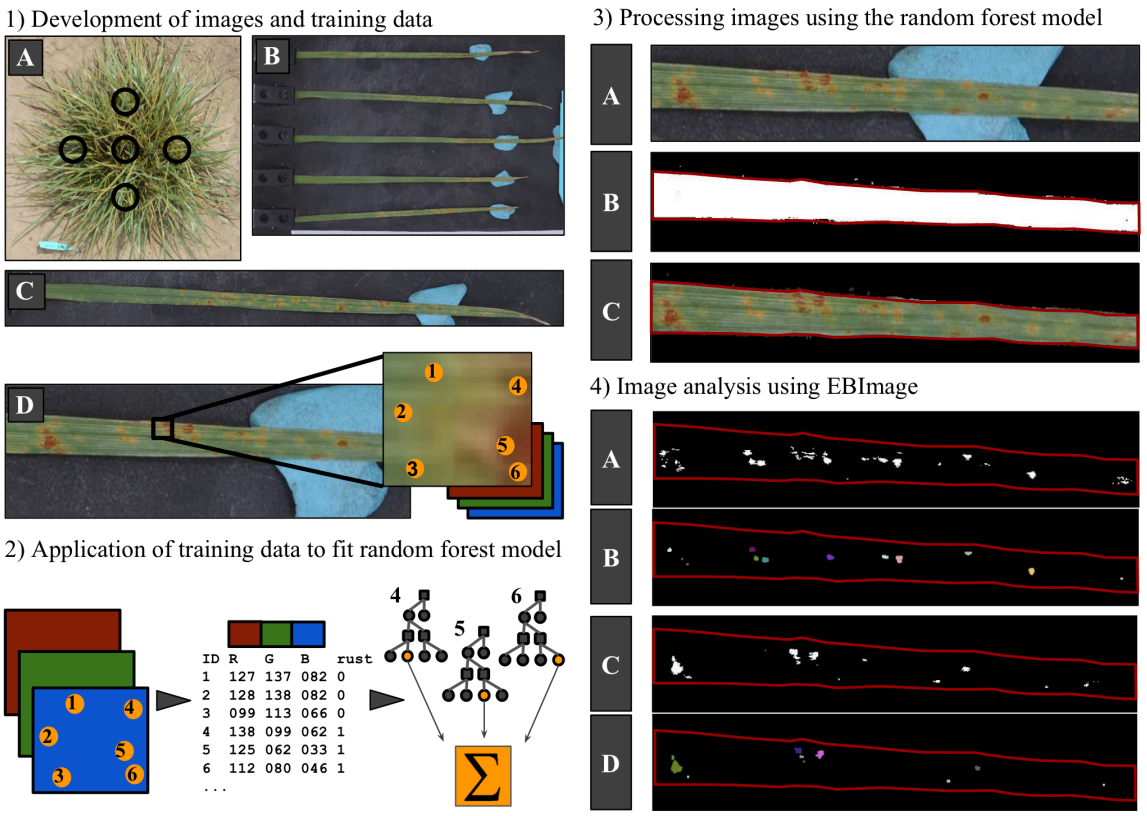


Fig. 1.1. Diagrammatic description of the image analysis pipeline: (1) examples of image series: (1A) whole-plant (WP) with leaf sampling pattern, (1B) five-leaf (FL), (1C) single-leaf (SL), and (1D) six example training data points to demonstrate the categorization of leaf pixels (Data Points 1–3) and rust pixels (Data Points 4–6); (2) random forest models were fit using the numeric RGB values associated with selected pixels (shown in 1D); (3) foreground is separated from background: (3A) random forest was applied to a cropped image, (3B) high-probability foreground pixels are colored white, (3C) EBImage was used to fill any gaps within the predicted foreground space, and coordinates of probable pixels from image 3A were overlaid; (4) crown and stem rust were separated from the foreground (3C): (4A) masked image of high-probability crown rust pixels, (4B) adaptive thresholding was applied to find circular objects in the mask, (4C) masked image of high-probability stem rust pixels, (4D) adaptive thresholding was applied to fill holes in large objects in the mask. Watershed transformation was used in 4B and 4D to separate and tag pustules in close proximity with different colors for quality control analysis.

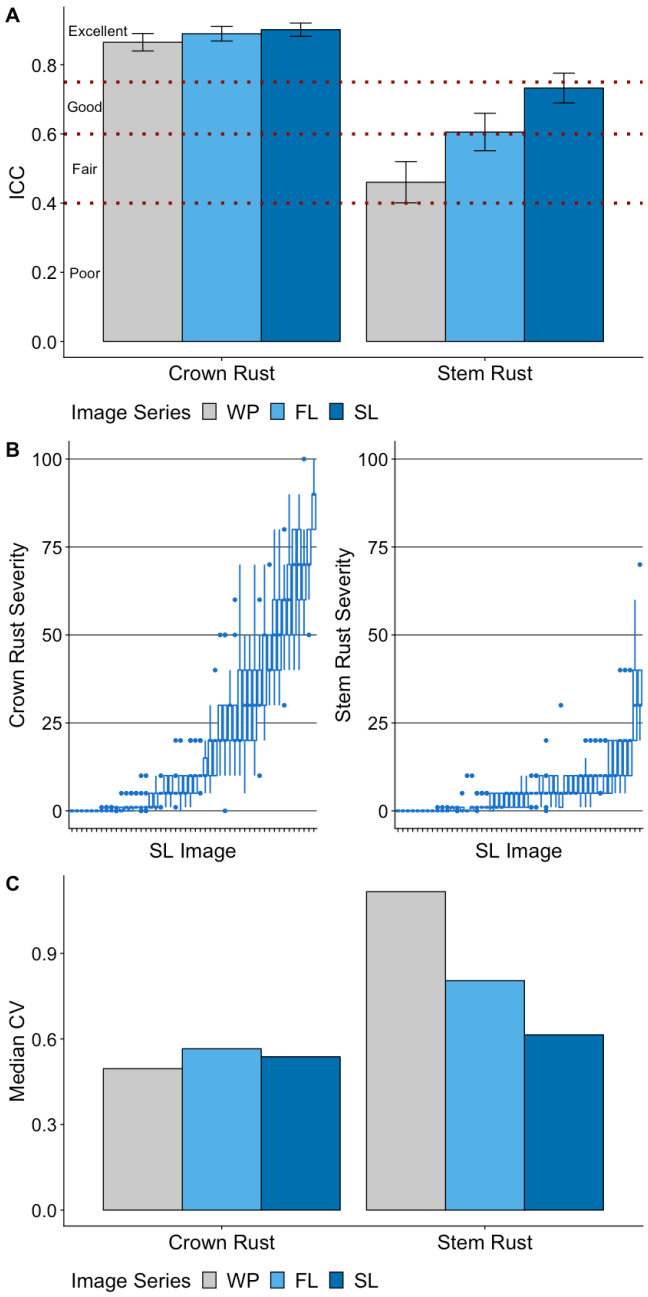


Figure 1.2 - Assessment of rater (n = 9) variability for crown and stem rust. A) Intra-class correlation coefficients (ICC) for both crown rust and stem rust across all three-image series. Higher ICC values indicate greater inter-rater reliability, or rust severity score consistency. Bars surrounding the coefficient are equivalent to +/- 1 standard error. Red dotted lines represent cut-offs for poor, fair, good, and excellent values. B) Rater variability described using boxplots for SL images only. Images were sorted by lowest to highest mean score. C) Median coefficient of variation (CV) value for each image series and disease.

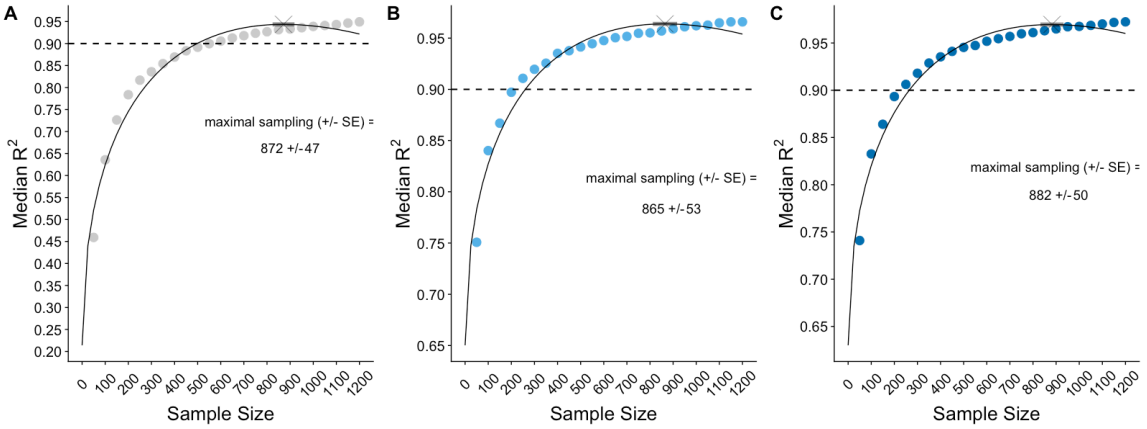


Figure 1.3 - Training data sample size requirement for each series described as the number needed to reach either an R^2 of 0.90 or the maximal sampling value. Median R^2 values are the median of 100 random forest iterations. A) WP images required 500 samples to reach R^2 of 0.90. B and C) FL and SL images required 250 samples to reach R^2 of 0.90.

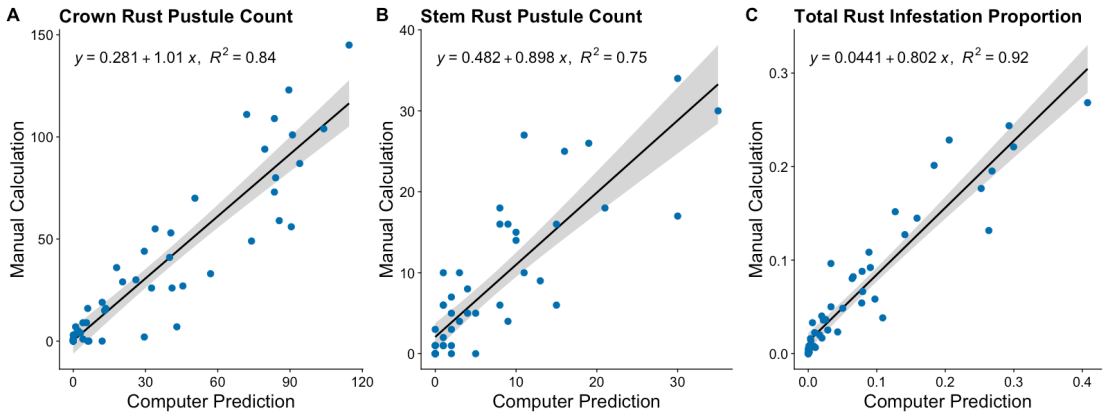
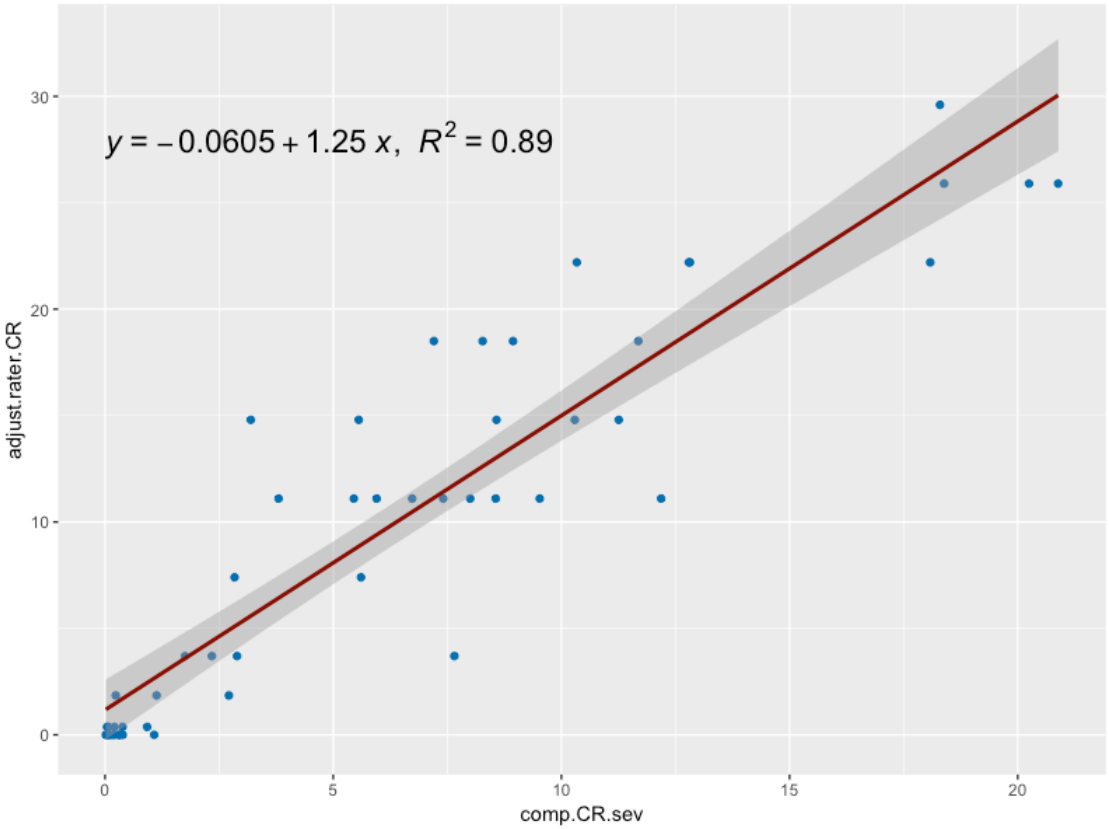
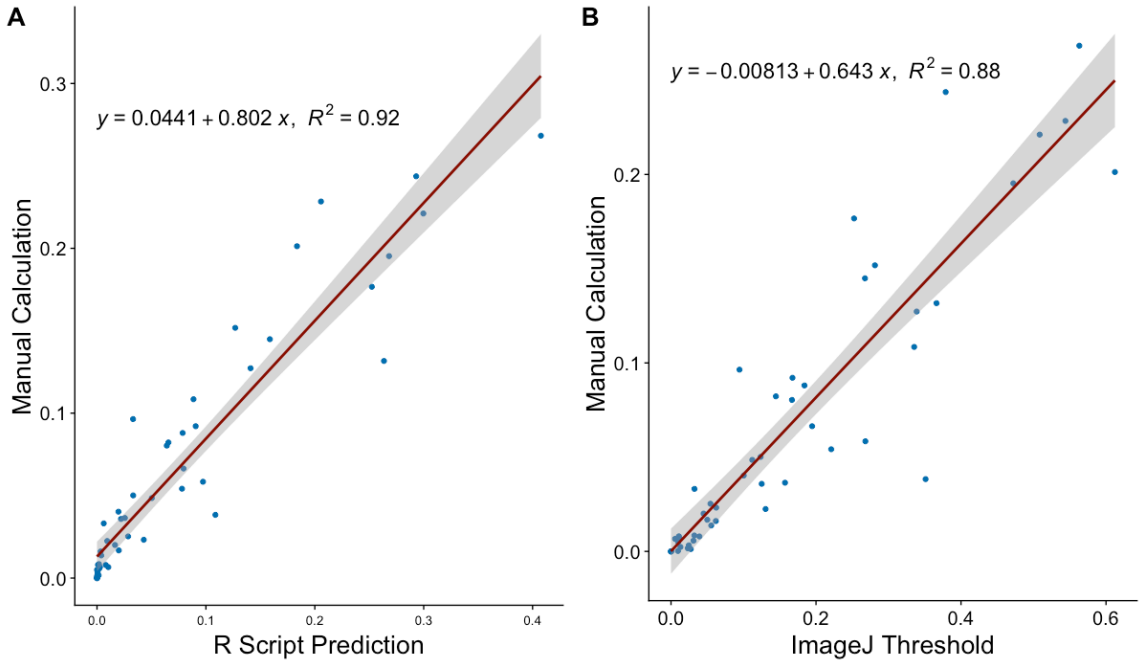


Figure 1.4 - Validation of computer accuracy for SL images. Single-leaf images were used because metrics such as pustule number and area could be easily quantified manually. The x axes in each panel are the computer value and the y-axes are the manually collected value. A) Crown rust pustule number. B) Stem rust pustule number. C) Proportion of leaf infested with rust (crown + stem). Shaded area represents a 95% confidence interval around the regression line.

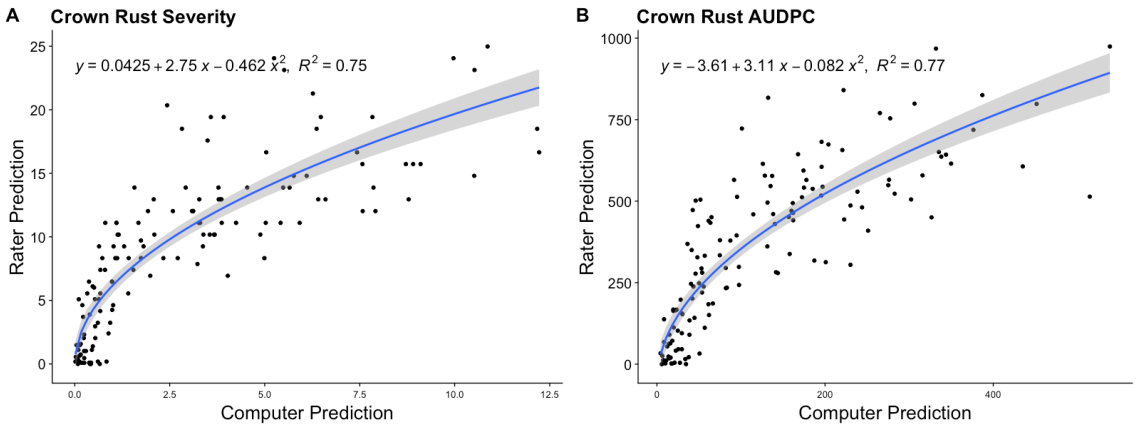
Supplementary Figures



Supplemental Figure 1.1- Linear regression between computer and median rater scores for crown rust on whole plants. Model coefficients show that computer scores are generally lower than the median rater score.



Supplemental Figure 1.2 - Total rust severity as calculated in ImageJ and R on single leaves. Severity was calculated by manually drawing polygons around each crown and stem rust pustule (y-axis) and by A) using total severity calculated in R and B) manually thresholding each image in ImageJ based on approximate HSV parameters.



Supplemental Figure 1.3 – Results from the validation experiment that included both a single crown rust severity rating and the area under the disease progress curve from five ratings. The validation data set included all 560 plants in the nursery from which the 50 plants in this study were selected from. A) Severity was estimated visually in the field by a single rater and regressed against image analysis scores from images taken on the same day. B) Area under the disease progress curve calculated based on five visual scores and five image series taken over 45 days in 2017. Data shown in A and B were best modeled with a quadratic function.

Chapter 2

The fungal endophyte *Epichloë festucae* var. *lolii* plays a limited role in mediating crown rust severity in perennial ryegrass

Introduction

Perennial ryegrass (Lolium perenne L.) has been bred as both a forage and turfgrass since the mid 20th century. Currently, there are over 300 registered cultivars of perennial ryegrass in the US and at least 80,000 metric tons of certified seed sold in Europe and the US annually (Sampoux et al., 2013; AMS, 2018; NASS, 2018). Crown rust, caused by the biotrophic fungal pathogen Puccinia coronata f. sp. lolii, commonly infects perennial ryegrass regardless of growing region and drastically reduces the value of susceptible cultivars by decreasing dry matter up to 37% and green tissue by 94% (Clarke and Eagling, 1994). These losses negatively impact the forage and turfgrass industries by reducing the yield, nutritional quality, and aesthetics of stands (Smit et al., 2005).

Increasing resistance to crown rust through breeding has led to measurable gains; however, P. coronata commonly overwinters on buckthorn (Rhamnus cathartica L.) and reproduces sexually allowing for yearly recombination that rapidly increases pathotype diversity (Sampoux et al., 2013). A study by Schubiger and Boller (2016) exemplified this diversity by identifying 107 pathotypes from 112 total isolates of P. coronata, of which at least two were virulent on each individual in a differential set of 32 perennial ryegrass grass hosts. Perennial ryegrass cultivars are usually diverse populations derived from polycrossing many parents, which makes incorporating specific resistance genes difficult due to segregation in the progeny. Practitioners may control crown rust through the use of fungicides; however, this may not be an option depending on the country the grass is grown and can be detrimental to the environment (Tilman et al., 2002). Recently, controlling disease through the use of biological controls such as Trichoderma species, mycorrhizae, and endophytes has been studied as an alternative (Ghorbanpour et al., 2017).

Perennial ryegrass, like many cool-season grasses, is often associated with a Clavicipitaceous endophyte (Leuchtmann et al., 2014). Epichloë festucae var. lolii naturally inhabits perennial ryegrass and has been frequently studied for the capacity to reduce animal herbivory (Clay, 1989). Research has shown that a single grass plant usually can only support a single Epichloë strain (Wille et al., 1999; Shoji et al., 2015); however, substantial diversity between isolates from different hosts has facilitated the development of commercial endophytes, which have been successfully integrated into cultivars producing an estimated revenue of $200 million in New Zealand (Johnson et al., 2013).

Unlike biological controls such as Trichoderma spp. that need to be inoculated, E. festucae var. lolii grows intercellularly and is vertically transmitted from mother plant to seed with the ability to remain viable for many years if seed is stored properly (Bylin et al., 2016). Research has demonstrated that endophytes can have positive effects on several abiotic and biotic stresses and show promise against several pathogens both in controlled environments and in the field (Xia et al., 2018). To date, many studies have examined the effect of Epichloë spp. on rust incidence in the Pooideae. Christensen and Latch (1991) studied the in vitro effects of E. coenophiala on germination and growth of stem rust (Puccinia graminis subsp. graminicola) urediniospores and found that three of thirty endophyte isolates showed an inhibitory effect. Conversely, a field study by Welty et al. (1991) found E. coenophiala had no effect on stem rust severity in tall fescue (Schedonorus arundinaceus [Schreb.] Dumort.). Two years later, Welty et al. (1993) conducted a similar study in the tall fescue cultivar 'Kentucky 31' and found initial signs of reduced stem rust incidence but detected no significant differences as disease severity increased. These findings suggest limited impact of the endophyte on stem rust resistance in the field, at least in tall fescue. Clay (1989) offered anecdotal evidence for increased rust resistance from infection with E. festucae var. lolii in perennial ryegrass, but this was not empirically tested. Ravel et al. (1995) conducted a field study and found no significant effect of endophyte on crown rust severity in perennial ryegrass across several locations and host families. Similar results were also found for meadow fescue (Festuca pratensis Huds.) infected with Neotyphodium uncinatum against Puccinia spp. (Pańka et al., 2011).

Despite the majority of papers reporting that endophytes do not impact rust severity, several literature reviews strongly suggest that endophytes could improve resistance to several rust diseases (Clay and Schardl, 2002; Kuldau and Bacon, 2008; Xia et al., 2018). Evidence to support these conclusions have generally stemmed from inhibitory effects determined through in vitro testing, preliminary observations, or grey literature (Ford and Kirkpatrick, 1989; Yue et al., 2000; Panka et al., 2004). However, proposed mechanisms by which endophytes could mediate general disease resistance are logical and include secretion of antifungal compounds, competition for intercellular and exterior space through epiphyllous hyphal nets, and modification of litter dwelling communities (Becker et al., 2016; Perez et al. 2017; Tian et al., 2017). Although these mechanisms are based on sound

theory, direct, correlative, field-based evidence has yet to be published. Rust epidemics progress rapidly in the field under favorable conditions and therefore rating is often conducted as a repeated measure and quantified using the modified Cobb scale and area under the disease progress curve (AUDPC) (Roelfs, 1992; Van Der Plank, 1963). Visual ratings have long proven effective; however, data taken by several raters on the same experimental unit over time have been shown to be imprecise for several foliar pathogens including Puccinia spp. (Sherwood et al., 1983; Heineck et al., 2019). Therefore, quantitative ratings using image analysis can be used to reinforce visual ratings in the field.

Unanimity surrounding endophyte impact on specific diseases has likely been clouded because of the difficulty in separating host effects per se from endophytic effects as environmental conditions often interact with both biological constituents. Adding further confusion is the strain diversity found within and among plant germplasm collections (Reed et al., 2000; Hettiarachchige et al., 2015). Isolating the exact effect of an endophyte on an individual or population of hosts has been done through the use of several experimental designs. Endophyte infected (E+) and endophyte free (E-) hosts within an experiment can be randomly selected from a population or cultivar (non-isogenic), identical (isogenic), or related by maternal background (isofrequent) (Heineck et al., 2018). Isogenic, or genetically identical, populations or genotypes are commonly used owing to any effect of host genotype being theoretically removed. Endophyte removal from a host can be accomplished by using systemic fungicides such as benomyl or propiconazole (Hesse et al., 2003; Kane, 2011). Isofrequent populations are the progeny of isogenic, or closely related, parents and are congenitally endophyte infected or endophyte free, based on parental infection status. Therefore, crosses between E+ and E- individuals result in similar allele frequencies between the resulting E+ and E- families (Casler and van Santen, 2008). Several studies have used isogenic or isofrequent populations to study the effect of endophyte on disease severity (Bonos et al. 2005, Perez et al. 2017). Using both isogenic and isofrequent designs can be especially useful for breeding purposes as endophyte effects after recombination can be assessed (heritability). In the context of rust resistance, it may be useful to compare the heritability of host rust resistance to the stability of endophyte effect across generations.

The objective of this study was to observe diverse E+ and E- isogenic and isofrequent perennial ryegrass populations to study the effect of endophyte infection on crown rust severity in the field.

Materials and Methods

Germplasm Selection

Experimental entries consisted of 14 perennial ryegrass selections, 3 of which were cultivars and 11 of which were wild or landrace accessions (Fig. 2.1). Seed from several entries originated from the same country (entries A, C, and D) but were tested separately because differences in collection location could lead to different host and endophyte effects. The cultivar 'NK200' is a forage variety, while 'Green Emperor' and 'GrandSlam GLD' are turf-type cultivars developed by the University of Minnesota and Peak Plant Genetics, respectively. Each entry tested positive for endophyte infection before experimental populations were designed.

Endophyte Detection and Isolation

Plant material was first tested for endophyte infection by tissue print immunoblot using a commercial kit (Phytoscreen Immunoblot Kit ENDO797-3, Agrinostics, Watkinsville, Georgia, USA, http://www.agrinostics.com) (Hiatt *et al*., 1997). Scoring was done on a binomial scale, with any ambiguous blots resulting in the plant material being removed from the experimental design. Additional screening was also conducted to confirm immunoblot results using light microscopy with methods described by Florea *et al.* (2015).

In order to identify the species, diversity and phenotype of each isolate, the Clavicipitaceous endophyte from each selected host was isolated using the methods described by Florea *et al*. (2015). Initial isolations were grown on PDA (full strength potato dextrose agar) media with 100 µg/mL ampicillin. Mycelium was sub-cultured after two weeks of growth onto fresh full-strength PDA plates without ampicillin. Cultures were then incubated in the dark at 23°C and sub-cultured when necessary.

Population Design

Isogenic: Five E+ plants were selected from each of the 14 entries. These 70 genotypes constituted the host genetics of the isogenic populations. Isogenic E+ and E- pairs were made by vegetatively propagating each of the 70 E+ genotypes and then subjecting one of the resulting clones to five full-rate foliar applications of propiconazole (Kestrel MEX, Phoenix Environmental Care, LLC). This resulted in two identical clones for each perennial ryegrass genotype: one E+ the other E-. Endophyte status was checked 60 days post-application to confirm endophyte knockout (Hiatt *et al*., 1997).

Isofrequent: The five isogenic E+ and E- pairs within each entry were the progenitors of five isofrequent half-sib family pairs. Each isogenic parent was vernalized for 114 days in a walk-in cooler set at 3°C under 10 h d^{-1}. Then, the host genotype pairs within each accession were brought into the greenhouse to be polycrossed in isolation from any other flowering *Lolium* or *Festuca* spp. The isolated cross for each accession resulted in half-sib seed from each isogenic pair. Each E+ and E- family consisted of precisely the same maternal background and similar paternal allele frequencies. For a more detailed explanation of this type of population design, see Heineck *et al*. (2018). Endophyte status of six random half-siblings from each of the selected isofrequent pairs was tested before the experiment started (Hiatt *et al*., 1997).

Experiment 1: Isogenic Populations

Experiment 1 was conducted in both 2017 and 2018 at a different field site each year. The same 70 isogenic pairs (140 treatments) were trialed each year and were propagated from the same stock material. The experiment was a 14 x 5 x 2 nested design with 5 random genotypes nested within each of the 14 entries. Each genotype was represented by an isogenic E+ and E- pair. Plants were arranged in a randomized complete block design with four replicates (560 experimental units each year). This design was used to analyze the effect and interaction between the endophyte and its host at both the genotypic and entry level.

Experiment 2: Isofrequent Families

Experiment 2 was conducted only in 2018. One isofrequent family pair was chosen from each accession, excluding entry D due to insufficient seed production. The 13

remaining family pairs were selected based on any preliminary findings of endophyte effect from visual ratings in 2017 ($\alpha = 0.10$); if no effect was found, an isofrequent pair was randomly selected. Each isofrequent family pair was represented by 24 genotypes. The experiment was a 13 x 2 factorial with 13 families with and without endophyte arranged in a randomized complete block with 12 replicates. This design was directed at confirming endophyte effects found in Experiment 1 by observing interactions at the host family level.

Field Conditions and Management

Individuals for each experiment were grown and maintained in a greenhouse and then transplanted into the field in May. Experimental trials were planted at a different location at the University of Minnesota Agricultural Experiment Station in St. Paul, MN in 2017 and 2018. Each experimental unit was observed in a spaced plant environment with plants 0.61 m apart. The crown rust susceptible cultivar 'Linn' was planted around each nursery to help spread natural inoculum. Weeds were controlled by hand and with a single application each year of 0.8 L ha^{-1} 2, 4-dichlorophenoxyacetic acid (Shredder, Winfield Solutions) and 1.5 L ha-1 dithiopyr (Dimension 2EC, Dow AgroSciences). Starter fertilizer was applied at a rate of 37 kg N ha^{-1}, 61 kg P ha^{-1}, and 37 kg K ha^{-1}. Supplemental irrigation was added as needed during establishment using a Kifco ST3 irrigator (Kifco, Havana, IL).

Data Collection

Data was collected on rust severity by both visual ratings and digital image analysis. Visual assessment of crown rust was done using the modified Cobb scale (Roelfs, 1992). Images and visual measurements were taken as repeated measures for both experiments. Experiments 1 and 2 each had five visual ratings conducted with ratings being completed about seven days apart. The first rating was conducted before the onset of disease. Spaced plant imaging for Experiment 1 was conducted six times in 2017 and seven times in 2018. Similar to visual ratings, imaging began before the onset of disease and series were captured about seven days apart. Five-leaf samples from each plant in Experiment 1 were collected once each year when rust severity was at its greatest. Plants in Experiment 2 were imaged once.

Image Analysis for Crown Rust

Images were taken using Nikon D300 digital SLR camera and saved as high-quality JPEG files of 4288 x 2848 pixels. Imaging was conducted under ambient conditions in 2017 between 1200 and 1400 hours with a fixed monopod mounted 80 cm above the ground, squarely above the plant. Camera settings were set to a shutter speed of 1/160 s, aperture setting of F8, white balance 5000K, and a focal length of 80 mm. In 2018 images were taken using the same camera but with a light box. The light box was constructed of an aluminum frame (0.8 m wide x 0.6 m long x 0.9 m tall) with coroplast walls lined with mylar to eliminate light infiltration. The interior was lit by 200 cm of high color rendering index LED light strip with an output of 221 lumens per 0.09 m^2 and a 4000K color value (superbrighLED.com). The lights were powered by a rechargeable TalentCell 12-volt lithium ion battery pack (talentcell.com). Camera settings for light box images were shutter speed 1/80 s, aperture setting 2.7, white balance 5000K, and a focal length of 80 mm.

The leaf samples from each were collected in a cross pattern on each plant to limit sample bias. Each leaf was removed from the plant where the collar meets the pseudostem, so the entire lamina was imaged. Leaves that had not fully emerged from the sheath, had not developed a visible collar, or were completely necrotic were not sampled. Leaf samples were imaged in the field immediately after placement on a matte black imaging board. The leaves were imaged in a box lit by 150 cm of high color rendering index LED light strip with an output of 221 lumens per 0.09 m^2 and a 4000K color value (superbrighLED.com). The lights were powered by a rechargeable Talentcell 12-volt lithium ion battery pack (talentcell.com).

Quantification of rust severity was conducted using an automated system described by Heineck et al. (2019) for both five-leaf samples and whole plants. Image analysis was done using ImageJ and R (Schindelin *et al.,* 2012; Version 3.5.0; R Core Team, 2018). Core packages in R were *EBImage* and *randomForest* (Liaw and Wiener, 2002; Pau *et al*., 2010). Random forest models were generated for both leaf samples and whole plant images. Separate training data sets were made for each year using eight images across 2017 and 2018, from which 300 data points were collected to train the model. Computer training was directed on the crown rust pustules *per se* to limit bias from other diseases such as stem rust (*Puccinia graminis* subsp. *graminicola*). Image analysis output at each time point

was inspected visually for any abnormalities in the processing pipeline. The proportion of infected tissue with crown rust was collected as a repeated measure on whole plant images. Severity was calculated as the quotient of probable crown rust pixels to healthy leaf pixels times 100. Leaf area (cm^2), crown rust pustules $leaf^{-1}$, and average pustule size (cm^2) was measured from leaf plate images. Area was calculated based off of a ruler that was included in each image by measuring the number of pixels in 1 cm line.

Identification of Fungal Endophytes

To identify endophytes isolated from this study, one isolate was randomly selected to represent each entry. Additional endophyte isolates were selected on the basis of consistent endophyte effect on crown rust in Experiments 1 and 2. Endophyte identification was accomplished using the approach described by Giauque *et al*. (2018). Briefly, the internal transcribed spacer (ITS) and partial large subunit (LSU) sequence regions of each isolate were sequenced. Sequences were blasted to the NCBI database using BLASTn for endophyte identification. Details of endophyte identification are available (SI Methods S1).

To investigate the diversity of the sequenced endophyte isolates, LSU and ITS sequences were concatenated and aligned using MAFFT (Katoh and Standley, 2013). Alignment was inspected and regions with poor alignment were trimmed using trimAI online server (Capella-Gutiérrez et al., 2009). Maximum likelihood estimation was calculated using RAxML under GTR+G model with a 1000 iteration bootstrap. A phylogenetic tree was visualized using FigTree (v 1.4.3) (Stamatakis, 2014; Rambaut, 2012). Sequence pairwise comparison was also performed using NCBI BLASTn (v 2.9.0) and visualized in R. All sequences generated from this study were deposited in NCBI GenBank (Table 2.1). Finally, phenotypic data were collected on each of the selected isolates according to Christensen and Latch (1991). Traits included growth rate, color, shape, and texture; these were observed on mycelia growing on full strength PDA media with 100 µg/mL ampicillin added.

Statistical Analysis

The repeated measures of crown rust severity by visual and computer estimation were used to calculate AUDPC. R code was adapted from published sources and R package

'agricolae' to automate the calculation of AUDPC for each experimental unit for both experiments (Madden *et al.,* 2007; Sparks *et al.,* 2008; Mendiburu, 2016).

Area under the disease progress curve values were not normally distributed and required transformation prior to ANOVA to meet assumptions of normality. Visual severity data required log(x+1) or square root transformation. Image data were transformed using a logarithmic transformation. Data were analyzed using a linear mixed effects model with R package 'lme4' (Bates *et al*., 2015). For Experiment 1, entry, host genotype, and endophyte were considered fixed effects with block being the only random effect. Host genotype was nested within entry. Models for Experiment 2 assigned half-sib family and endophyte as fixed effects and genotype as a random effect. Some data were unbalanced due to plant death, therefore R package 'emmeans' was employed to predict marginal means and calculate standard error (Lenth, 2018). All data were back-transformed prior to plotting estimated marginal means. All results were plotted using R package 'ggplot2' (Wickham, 2016). Pairwise comparisons were done on isogenic and isofrequent pairs within entry using a Bonferroni *P*-value correction ($\alpha = 0.10$). Broad-sense heritability was calculated on a genotype mean basis across environments for Experiment 1 using a fixed effects model and nesting blocks within years. The equation provided demonstrates how variance was partitioned to estimate heritability:

$$H^2 = \frac{{\sigma_G}^2}{\frac{{\sigma e}^2}{ry} + \frac{{\sigma_{GE}}^2}{y} + {\sigma_G}^2}$$

Where r = reps and y = years, σ_G^2 represents genotypic variance, and σ_{GE}^2 is variance for genotype x environment interaction. Narrow-sense heritability was calculated for Experiment 2 by way of parent-offspring repression. The maternal parental and progeny AUDPC mean values were standardized and then regressed with the resulting slope coefficient equal to narrow-sense heritability. The unknown paternal parent was considered random.

Results

Experiment 1: Visual vs. Computer Rating

The correlation coefficients between visual and computer AUDPC values in Experiment 1 were robust in both 2017 and 2018 (r = 0.88 and r = 0.87, $P < 0.001$)

(Supplemental Fig. 2.1). In general, visual ratings had much higher AUDPC values compared to computer predictions. This was simply due to differences in the Cobb scale itself compared to actual percent infestation on the leaf (x*0.37), which the computer measures. Coefficients of variation were similar for both computer (CV = 0.26) and visual (CV = 0.29) scoring methods.

Grass Host Effects

Analysis of variance for Experiment 1 showed a substantial main effect of entry and interaction between entry and host genotype across years (Table 2.1). For example, entry R (PI 610806) had almost no rust incidence in either 2017 or 2018. However, within many entries variation between genotypes for crown rust severity was large (Fig. 2.2). It is important to note that because many hosts were not susceptible to crown rust the effect of endophyte could not always be determined. Broad-sense heritability was calculated across years on a genotype within entry-mean basis and was substantial for both visual ratings (H^2 = 0.83) and computer ratings (H^2 = 0.76). In Experiment 2, there was a strong effect of family ($P < 0.001$) for both the visual and the computer rating. When standardized half-sib AUDPC values were regressed against mean maternal parental means, a highly significant and positive slope was found (β = 0.76, R^2 = 0.59, P = 0.002). The slope estimate (β) is equal to narrow-sense heritability. Mean isogenic parental and isofrequent offspring AUDPC rankings were also very comparable (r_s = 0.80, P = 0.002).

Experiment 1: AUDPC

Separate models were fit for 2017 and 2018 due to the different rating dates and number of assessments comprising the AUDPC values across years inappropriate. Analysis of variance found significant three-way interactions between entry, genotype, and endophyte for both years and scoring methods, which resulted in observing effects on an individual host genotypes. The F values for the three-way interaction were substantially less than those for entry and genotype (Table 2.1). High-order interactions aside, the main effect of endophyte was not consistently significant. Computer rating in 2017 found a negative main effect of endophyte on crown rust severity (E- vs. E+ AUDPC 23.3 and 26.8 units, respectively, $P < 0.001$) (Table 2.1). Although the three-way interaction was always

highly significant and led to finding 28 significant contrasts (Fig. 2.2), effects within a year on any isogenic pair was considered with caution due to the number of comparisons made. Therefore, the focus of this analysis was on trends across years and data collection methods.

Overall, very few consistent endophyte effects were found between isogenic pairs. Of the 70 pairs, only D.7 and Y.7 showed consistent differences (Fig. 2.3). Interestingly, the effect of endophyte on rust severity was favorable for Y.7, but antagonistic in D.5. Consistent favorable effects of endophyte were observed across years for F.5 and E.3, but significance depended on disease scoring method. In some instances, repeatable differences were observed, but AUDPC values were very low for both infected and uninfected clones. For example, although genotype G.2 was negatively affected by endophyte infection, there was very low rust severity observed on that genotype in general (Fig. 2.2).

Experiment 1: Leaf Samples

There was no significant main effect of endophyte for any variable measured on leaf samples. Overall, variance was large within treatment group for leaf samples making statistical differences rare between isogenic pairs. Lower variance within treatment groups was observed in 2017 compared to 2018. This was likely due to more favorable weather conditions offering higher quality images. Analysis of variance showed a strong three-way interaction ($P < 0.001$) between entry, genotype, and endophyte in 2017 for leaf area, crown rust severity, pustule size, and pustule number per leaf (Table 2.2). However, very few significant pairwise comparisons between isogenic pairs were found for any of these traits (Supplemental Fig. 2.2). In 2018, no three-way interactions were found; however, a moderate interaction between endophyte and entry was found for leaf area ($P = 0.01$).

Percent area infected with crown rust on the five-leaf samples was strongly correlated to whole plant AUDPC in 2017 ($r = 0.87$, $P < 0.001$), but not in 2018 ($r = 0.56$, $P < 0.001$). Endophyte infected Y.7 clones trended towards lower disease severity (E+ = 0.3, E- = 2.3 % severity; $P < 0.24$) and fewer pustules per leaf in 2017 (E+ = 3, E- = 9 pustules leaf $^{-1}$; $P < 0.07$) and smaller average pustule size was observed in 2018. Antagonistic effect of endophyte on AUDPC in D.7 may have been the cause of increased

pustule number in 2017 (E+ = 39, E- = 22 pustules leaf $^{-1}$; $P = 0.86$) and increase pustule size in 2018 (E+ = 0.0044, E- = 0.0022 cm^2; $P = 0.07$) of E+ clones.

Experiment 2

Although multiple visual ratings were recorded for Experiment 2, only a single image series was captured during peak rust infection. Compared to Experiment 1, there was a slightly lower correlation between the last visual rating and computer score ($r = 0.77$, $P < 0.001$). Analysis of variance for AUDPC values calculated from visual ratings found an effect of entry ($P < 0.001$), but not of endophyte ($P = 0.99$). Analysis of the single image series taken found a significant interaction between endophyte and entry ($P = 0.002$) (Table 2.3).

An antagonistic effect of endophyte was observed in isofrequent families B.8 and J.4 for computer predicted severity (Supplemental Fig. 2.3), but this interaction was not observed in the parents of these particular families. Isofrequent family pairs Y.7, E.3, and F.6 had statistically similar AUDPC values and computer calculated severity values despite their maternal parents showing consistent favorable effect of endophyte (Fig. 2.4).

Endophyte Identification and Diversity

Sequencing was conducted on highly conserved ribosomal regions of DNA. All endophytes isolated in this study were identified as *Epichloë festucae* var. *lolii*. Results from phylogenetic analysis found that there was a very low probability of difference between isolates (Fig. 2.5B). Sequence pairwise comparison showed high sequence similarity between all isolates (Supplemental Fig. 2.4). The mean sequence similarity was 99.76% with the lowest similarity between A.1.1 and D.1.7 being 99.14%.

Endophyte isolates that showed consistent effect within an isogenic pair were observed in tandem with an isolate infecting a host that showed no effect on rust severity. For example, D.6 and D.7 both were susceptible to rust (sequence similarity 99.14%), but endophyte infection only influenced rust severity on D.7 (Fig. 2.2). Phenotypically, there were substantial differences observed among isolates from different entries for growth rate, texture, and shape (Fig. 2.5A). Conversely, differences between isolates within entry were

not apparent, for instance, D.6 and D.7 both had generally slow growth rates and a lobed phenotype (Fig. 2.5C).

Discussion

The two experiments presented in this study were designed to assess the potential of native perennial ryegrass *Epichloë* endophyte to control crown rust. The main effect of endophyte was never favorable or consistently significant across analyses of variance for both experiments. Results from isogenic populations in Experiment 1 showed that although endophyte by entry interactions were pervasive (Table 2.1), there was no evidence that endophyte infection could consistently influence crown rust severity on a population scale. The effects of endophyte on crown rust were limited to individual host genotypes across the diverse panel of perennial ryegrass germplasm (Fig. 2.1). In fact, there was never more than one isogenic pair within a single entry that showed an effect of endophyte. For instance, Y.7 and Y.0 were both susceptible to crown rust, but only Y.7 showed an effect of endophyte. This was also observed for D.7 and D.6, except the endophyte had an antagonistic effect on D.7 where it had a favorable effect in Y.7 (Fig. 2.3). Several studies have found that endophyte effects are not omnipresent across environments and are often highly specific, alluding to the complexity of host by microbe by pathogen interactions (Niones and Takemoto, 2014; Wiewióra *et al*., 2015).

In contrast to biotrophic fungal pathogens such as *Puccinia* spp., more pronounced effects are often observed when the causal plant pathogen is necrotrophic. For instance, tall fescue seedlings infected with *E. coenophiala* showed an increase probability of survival when challenged with *Rhizoctonia zeae*, the causal pathogen of damping off (Gwinn and Gavin, 1992). Red thread (*Laetisaria fuciformis*) had reduced virulence against several fine fescue species infected with the endophyte *E. festucae* (Bonos *et al*., 2005). Other examples include fine fescue infected with *E. festucae* showing an increased resistance to dollar spot (*Sclerotinia homoeocarpa)* and perennial ryegrass infected with *E. festucae* var. *lolii* with increased disease resistance to *Fusarium poae* (Clarke *et al*., 2006; Pańka *et al*., 2013). Pańka *et al.* (2011) studied the influence of *N. uncinatum* on the virulence of two biotrophic pathogens (*Puccinia* spp. and *Blumeria graminis*) and one necrotrophic pathogen complex (*Bipolaris sorokiniana*). The endophyte had no significant effect on either biotroph but did

reduce the leaf spot caused by the necrotrophic fungus. In the current study no effect of endophyte was observed across isogenic populations among or within entries, however individual isogenic pairs were explored.

The consistent three-way interactions in Experiment 1 between endophyte and genotype led to the discovery of several consistent effects of endophyte on AUDPC across years and measuring techniques. Consistent effect of endophyte on crown rust severity was observed for D.7 and Y.7 (Fig. 2.3). Notable effects were also found for genotypes G.2, F.6, and E.3 (Fig. 2.2). Of these five genotypes Y.7, F.6, and E.3 showed favorable effects. Image analysis of five-leaf samples provided measures of pustule number and size for each experimental unit. However, this information did not provide significant mechanistic evidence as to why endophytes had a significant impact on AUDPC measured on whole plants. Based on F-test values, endophyte infection had almost no effect compared to the effect of host (Table 2.2). Trends did suggest that E+ clones of D.7 had larger pustules compared to E- clones, while E+ clones of Y.7 had about 67% fewer pustules per leaf compared the E- clones. Most reviews indicate that endophytes have positive, or at least neutral, impact on disease severity (Clay and Schardl, 2002; Kuldau and Bacon, 2008; Bacon and White, 2016; Xia *et al.,* 2018). However, there are some opposing opinions; for instance, Bastias *et al.* (2017) made the argument that endophyte infection may downregulate immune responses such as salicylic acid-mediated immunity. This makes a convincing argument for why endophytes may actually reduce the ability of plant hosts to resist disease. Certainly, variability across isolates of *Epichloë* spp. has shown to influence alkaloid production, that in turn, influences the capacity to mitigate animal herbivory.

Then it is, of course, possible that within a single-entry genetic variation for *E. festucae* var. *lolii* is driving the favorable and antagonistic differences. These differences could explain why isogenic pairs of genotypes, such as Y.7 and Y.0, performed differently. Our investigation of mycelia phenotypes showed that growth rate and morphology were different among isolates from different entries; however, there were few differences within the same entry (Fig. 2.5A and C). Further Sanger sequencing confirmed all endophyte isolates were *E. festucae* var. *lolii* and isolates from the same entry usually group together (Fig. 2.5B), However, the sequencing was conducted on relatively conserved ribosomal DNA regions. Because of this, phylogenetic relationship inferred from the dataset might

not necessarily reflect the actual genetic variation between fungal isolates. Future genome sequencing on the isolates of interests could provide us insights on unique morphology characters we observed. With this in mind, it is important to note that *E. festucae* var. *lolii* is vertically and asexually transmitted and so although perennial ryegrass cultivars may have several potential endophyte genotypes half-sib families will only have one.

Experiment 2 utilized isofrequent families to validate the effects observed on genotypes that exhibited a favorable effect: Y.7, F.6, and E.3. Other families were also included in this experiment to assess narrow-sense heritability of crown rust resistance for hosts. Analysis of variance for visual AUDPC values showed no effect of endophyte (P = 0.99) and no interaction with entry (P = 0.12) for isofrequent families. These findings demonstrate there is strong specificity between endophyte and host for mediating crown rust. As proof, each of the three E+ isogenic families must have the same endophyte as the mother plant (Y.7, F.6, and E.3) and the host plants for both E+ and E- families must also share at least 1/8 of their alleles by descent. The combined results from Experiments 1 and 2 thus demonstrate: 1) endophyte infection does not induce a marginal effect over a diverse range of germplasm nor does infection within a single entry constitute a change in rust infection (Fig. 2.2) and 2) even when an effect was found on an individual genotypic level, it was not measurable in the progeny (Fig. 2.4). Endophyte-mediated disease suppression was found from the *E. festucae* Rose City isolate in isofrequent populations of fine fescue for red thread and dollar spot disease (Bonos *et al*., 2005; Clarke *et al*., 2006).

Difficulty in pinpointing the exact influence of endophyte infection on any disease stems from the far-reaching effect of host-microbe interaction, possibly through host mediation or quorum sensing (Bacon and White, 2016). Laboratory experiments have identified mechanisms by which endophytes mediate disease such as secretion of anti-microbial compounds and physical epiphyllous hyphae barriers (Tian *et al*., 2017). Yet, sporadic success for many diseases across field trials suggests caution in making blanket statements concerning endophyte-mediated disease suppression. For instance, Moy et al. (2000) proposed epiphyllous nets as a potential barrier against pathogen attack in fine fescue. However, this structure was not observed on perennial ryegrass infected with native *E. festucae* var. *lolii.* When perennial ryegrass was inoculated with non-native *E. festucae* epiphyllous hyphae were detected suggesting utility for this endophyte host combination

(Becker *et al.*, 2016). Further study is recommended on the genotypes that show an impact of endophyte infection. A reasonable next step would be to inoculate both Y.0 - Y.7 and D.6 - D.7 E- clones with *E. festucae* isolate 'Rose City' (Johnson-Cicalese et al., 2000). *E. festucae* may be able to grow epiphyllous nets on these ryegrass hosts (Becker *et al.*, 2016), and Rose City in particular, produces unique antifungal compounds (Tian *et al*., 2017). Crossing these two pairs would yield isofrequent families congenitally E+ 'Rose City' and E+ native endophyte. Challenging these E+ and E- families with both biotrophic and necrotrophic pathogens would result in a very robust study. Hemibiotrophic pathogens would be interesting to study relative to endophyte influence because their lifestyle may switch from survival on either living or necrotic tissue. Wäli *et al.* (2006) measured the effect of endophyte on the severity of snow mold (*Typhula ishikariensis*) in meadow ryegrass (*L. pratense*). They found a negative effect of endophyte infection on pathogen growth in both a field and greenhouse environments. Conversely, Greulich *et al*. (1999) investigated *E. typhina* impact on the virulence of *Cladosporium phlei* and found that plants with endophyte showed very little purple leaf spot development.

Pragmatically, there is little utility in native endophyte-mediated protection from crown rust infection, at least in the environments tested. Although maintaining horizontal resistance to crown rust can be challenging, breeding has facilitated large improvements over time with average yearly gains estimated at 8.9% (Sampoux *et al*., 2013). In this study we found broad-sense heritability estimates for both visual and computer ratings ($H^2 = 0.83$ and 0.76, respectively). A narrow-sense heritability estimate of 0.76 was calculated, indicating robust breeding potential. Based on these findings, improving crown rust resistance through traditional breeding methods such as recurrent selection and polycross-progeny selection will likely be more successful in the short term than pursuing endophyte-mediated control.

Conclusions

Infection with *E. festucae* var. *lolii* resulted in no consistent marginal effects on AUDPC across 14 perennial ryegrass entries. Within entry there were sporadic and variable effects on isogenic pairs resulting in three favorable interactions. The specificity of the three favorable endophytes were further explored using isofrequent families, however E+

and E- families did not differ for crown rust severity. The strong host specificity indicates that this endophyte has highly limited utility for controlling crown rust in the study environment. In contrast, high heritability estimates were observed for crown rust severity indicating potential gains from selection leading to increased horizontal resistance. Future work on endophyte-mediated rust resistance in perennial ryegrass could explore mechanistic reasons for both the favorable and antagonistic effects found at the host genotype level and why effects were not stable across generations. However, with regards to the vastly superior impact of host genotype on resistance, both within and among entries, further research on endophyte-mediated crown rust resistance should be considered on an individual host-genotype basis and not extended to a population scale.

Tables

Table 2.1 - Analysis of variance for crown rust area under the disease progress curve (AUDPC) for Experiment 1. Experiment 1 was repeated with different clones of the same isogenic plants in 2018. Rater AUDPC was calculated from scores recorded on the modified Cobb scale. Computer AUDPC values were based on predicted percent tissue infected.

		2017 Experiment 1			
		rater AUDPC		computer AUDPC	
Source	df	F value	*P* value	F value	*P* value
Entry (E)	13	60.91	<0.001 ***	78.21	<0.001 ***
Endophyte (Endo)	1	0.09	0.7505 NS	10.79	0.0011 **
E : Genotype (G)	56	21.77	<0.001 ***	28.41	<0.001 ***
E : Endo	13	1.95	0.025 *	1.95	0.023 *
E : G : Endo	56	1.62	0.0049 **	2.03	<0.001 ***
		2018 Experiment 1			
Entry (E)	13	73.08	<0.001 ***	102.13	<0.001 ***
Endophyte (Endo)	1	0.26	0.608 NS	0.02	0.8944 NS
E : G	56	20.16	<0.001 ***	27.69	<0.001 ***
E : Endo	13	3.62	<0.001 ***	5.24	<0.001 ***
E : G : Endo	56	2.21	<0.001 ***	2.05	<0.001 ***
*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels.					

Table 2.2 – Experiment 1 analysis of variance for leaf area, crown rust severity, mean pustule size, and pustules leaf $^{-1}$. Data was collected from five-leaf samples in 2017 and 2018.

	2017 Experiment 1								
		Leaf area		Severity		Mean pustule area		Pustule leaf $^{-1}$	
	df †	F value	P value	F value	P value	F value	P value	F value	P value
Entry (E)	13	42.5	<0.001 ***	39.9	<0.001 ***	11.0	<0.001 ***	42.5	<0.001 ***
Endophyte (Endo)	1	1.3	0.247 NS	0.6	0.46 NS	0.2	0.6313 NS	0.0	0.9616 NS
E : Genotype (G)	56	19.4	<0.001 ***	15.0	<0.001 ***	4.5	<0.001 ***	19.3	<0.001 ***
E : Endo	13	1.7	0.061 NS	0.9	0.53 NS	1.0	0.424 NS	0.8	0.7152 NS
E : G : Endo	56	1.9	<0.001 ***	2.5	<0.001 ***	1.5	0.020 *	2.3	<0.001 ***
	2018 Experiment 1								
Entry (E)	13	24.6	<0.001 ***	5.3	<0.001 ***	4.0	0 ***	4.5	0 ***
Endophyte (Endo)	1	0.2	0.685 NS	0.2	0.671 NS	0.9	0.353 NS	0.1	0.835 NS
E : G	50	5.3	<0.001 ***	2.8	<0.001 ***	2.5	0 ***	3.5	0 ***
E : Endo	13	2.1	0.013 *	1.2	0.312 NS	1.9	0.032 *	1.8	0.044 *
E : G : Endo	50	1.2	0.216 NS	0.9	0.633 NS	1.5	NS	1.1	0.305 NS

*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels.

† Degrees of freedom are lower for 2018 due to plat death

Table 2.3 - Analysis of variance for crown rust rater AUDPC and computer calculated severity in half-sib families included in Experiment 2. Half-sib families were the progeny of A.4, B.8, C.1, E.3, F.6, G.7, H.1, J.4, O.4, P.8, R.2, W.2, and Y.7.

	Experiment 2				
		rater AUDPC		computer severity	
Source	df	F value	*P* value	F value	*P* value
Family (Fam)	12	18.37	<0.001 ***	8.81	<0.001 ***
Endophyte (Endo)	1	0.00	0.991 NS	1.85	0.174 NS
Fam : Endo	12	1.44	0.118 NS	2.72	0.002 **
*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels.					

Figures

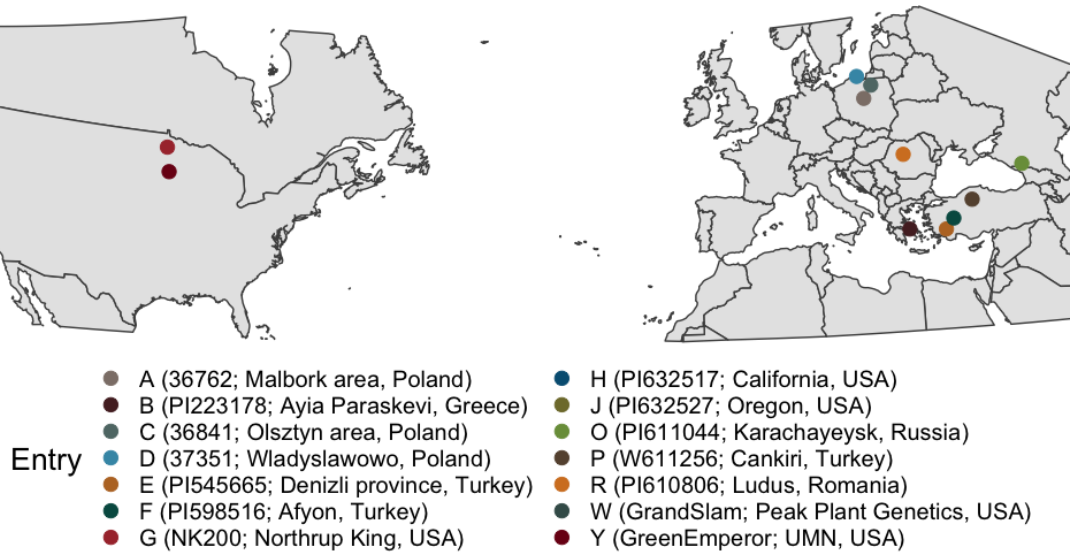


Figure 2.1 - Country and collection region of the 14 experimental entries. A coded letter designation was assigned to each of the accessions or cultivars. Entries G, W, and Y were commercial cultivars. Entries B, E, F, H, J, O, P, and R were obtained from the Germplasm Resource Information Network. Entries A, C, and D were landrace collections.

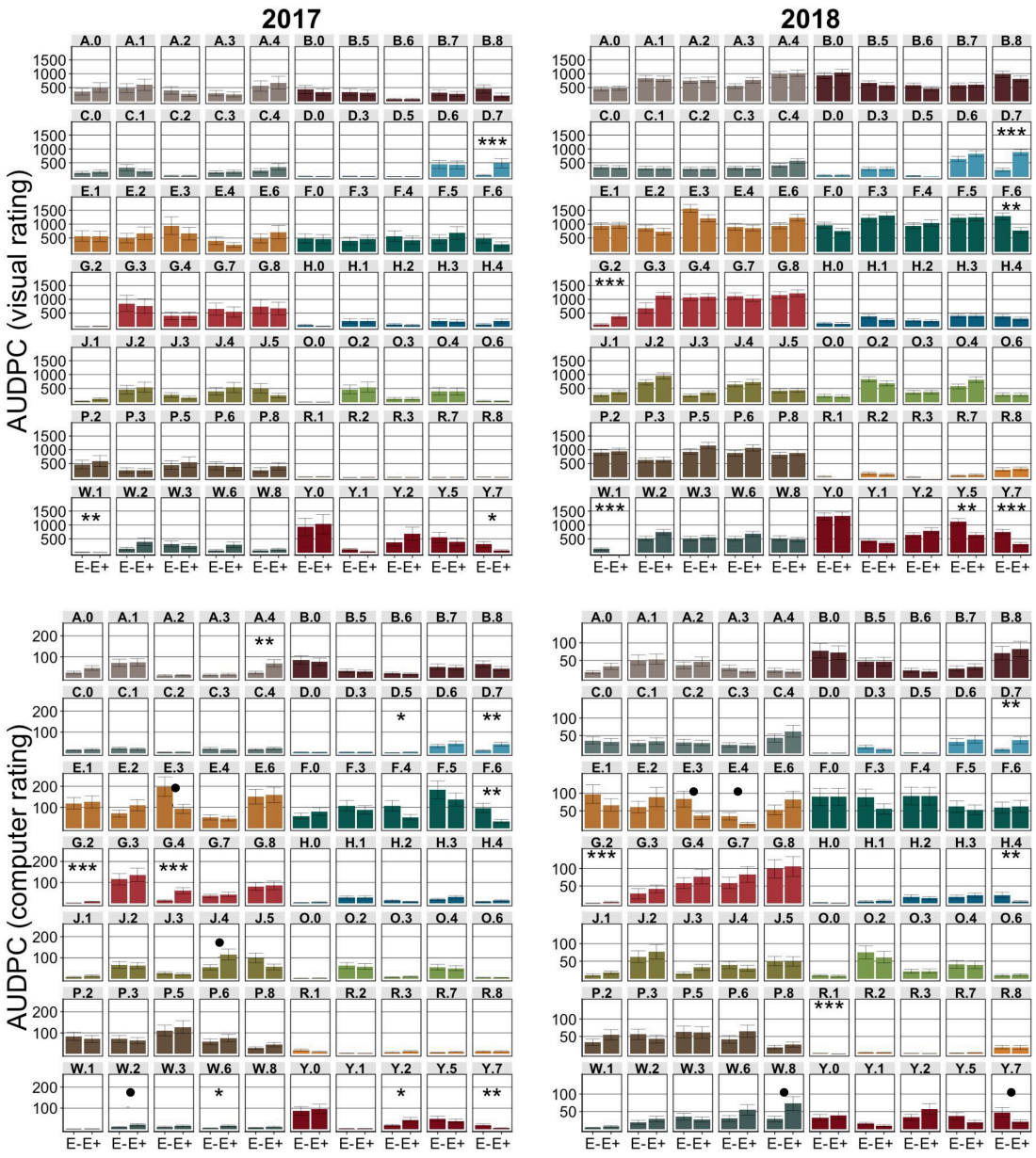


Figure 2.2 - Plotted estimated marginal AUDPC values for each isogenic pair in Experiment 1. Values in the top two quadrants were derived from visual estimates, whereas the bottom two were derived from computer predictions, with results presented from both 2017 and 2018. Each color represents an entry with the five genotypes nested within. Each window displays the estimated mean and significance for a single isogenic pair. Contrasts between pairs was conducted within entry (m = 5) with a Bonferroni P-value adjustment: '•' = 0.1, '*' = 0.05, '**' = 0.01, '***' < 0.001. Bars surrounding means are equal to one standard error. The large differences between visual and computer AUDPC values is due to the Cobb scale vs. actual percent severity.

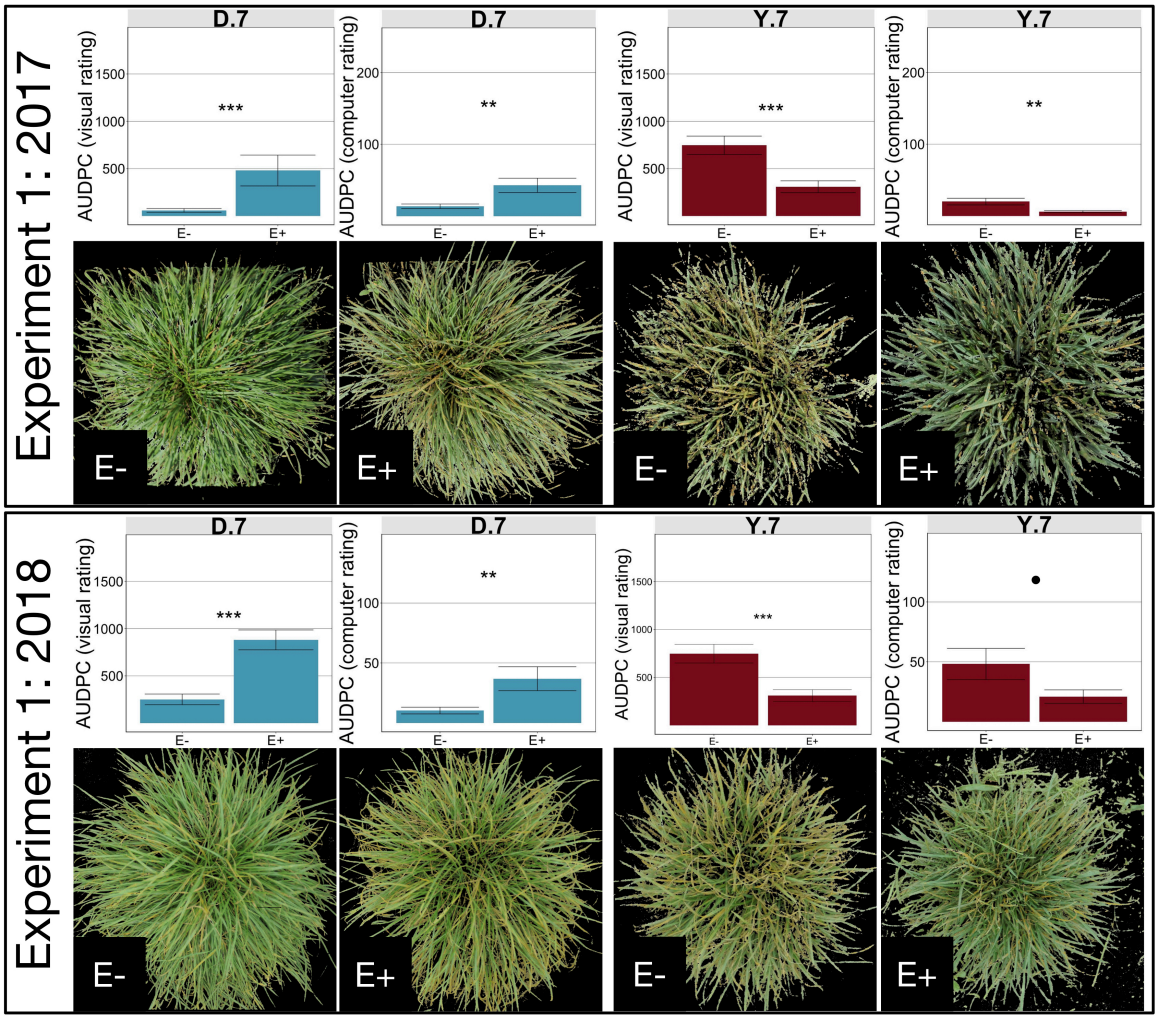


Figure 2.3 – Images of perennial ryegrass genotypes D.7 and Y.7 in Experiment 1. Displayed images were taken at peak rust severity in both 2017 and 2018. Area under the disease progress curves for both visual and computer scores are included with bars equal to one standard error. Contrasts between pairs was conducted within entry (m = 5) with a Bonferroni P-value adjustment: '•' = 0.1, '*' = 0.05, '**' = 0.01, '***' < 0.001.

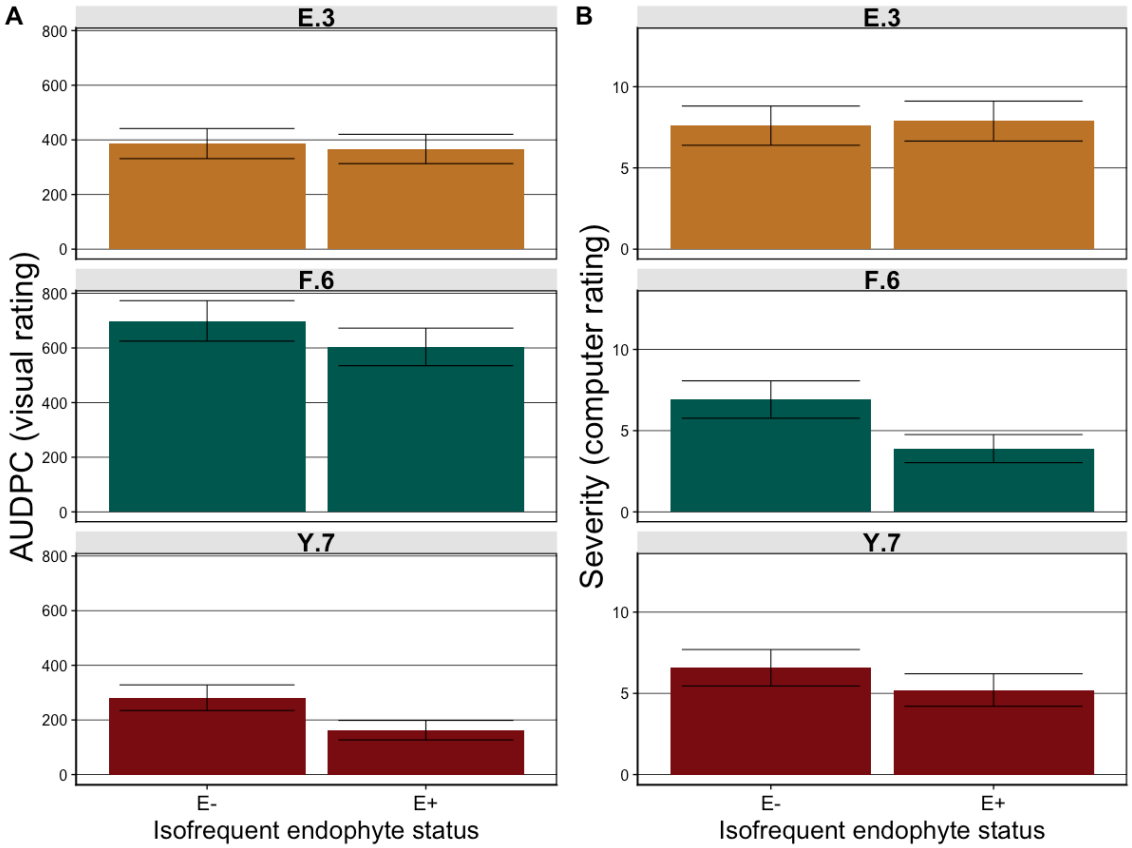


Figure 2.4 – Experiment 2 isofrequent families from isogenic pairs showing a favorable effect of endophyte. A) Plotted estimated marginal AUDPC values derived from visual scores. B) Plotted estimated marginal mean computer severity estimates in percent crown rust infection. Contrasts between pairs was conducted across entries with a Bonferroni *P*-value adjustment: '.' = 0.1, '*' = 0.05, '**' = 0.01, '***' < 0.001. Bars surrounding means are equal to one standard error.

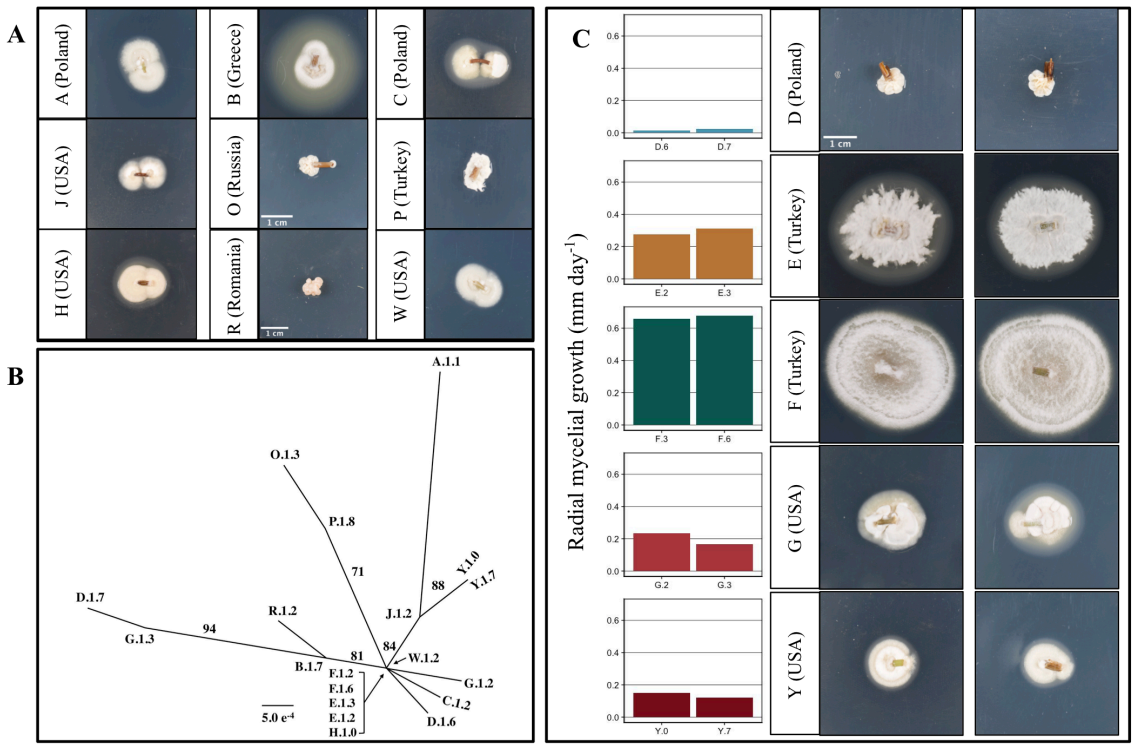


Figure 2.5 - A) Endophyte isolate phenotype after growing on full strength PDA for 40 days. B) Phylogenetic tree constructed by maximum likelihood (ML) method using concatenated LSU and ITS sequences generated from *Epichloë festucae* var. *lolii* isolated in this study. All endophytes shared high sequences similarity and were closely related, with some diversity across entries from different country of origin. Bootstrap percentages lower than 50 were not shown. C) Endophyte isolates from E+ hosts that showed effect on rust infectivity (Fig. 2.2) and a random isolate from the same entry that showed no effect. Endophyte growth was measured as mm of radial growth per day after 40 d.

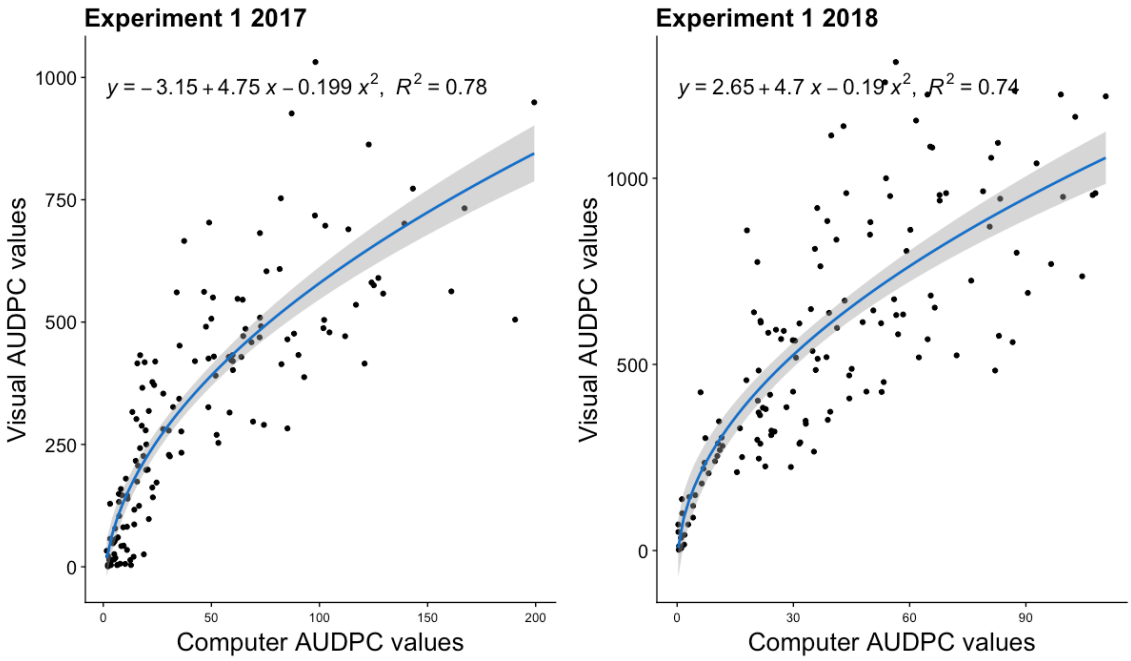
Supporting Information

Supplemental Methods 2.1: ITS and LSU region amplification

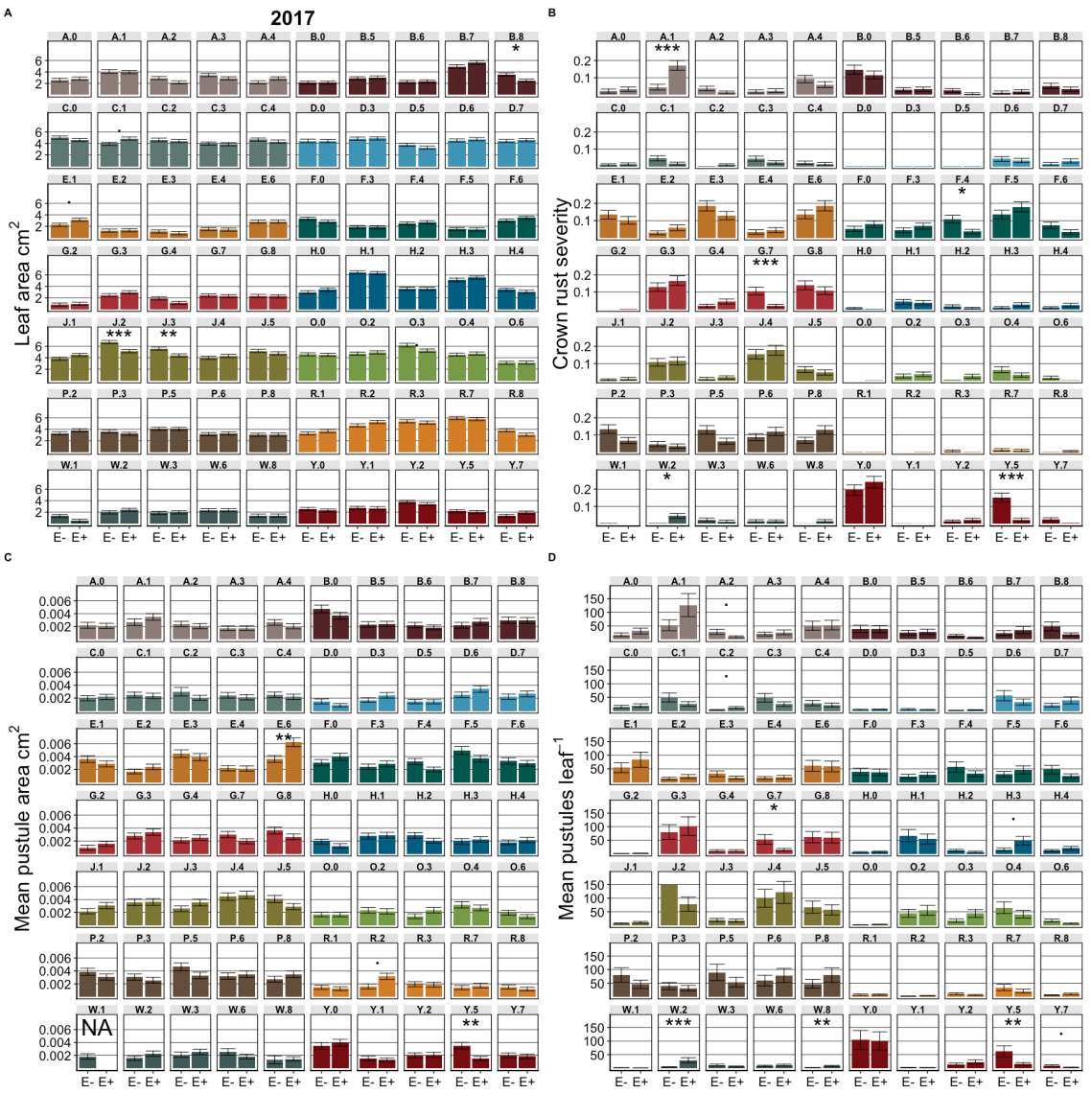
Genomic DNA of the isolates were extracted using Quick-DNA Fungal/Bacterial MicroPrep Kit (ZYMO Research) by following the manufacturer's instructions. Universal fungal primers ITS1 (5′-TCC GTA GGT GAA CCT GCG G-3′) and ITS4 (5′-GCA TAT CAA TAA GCG GAG GA-3′) were used to amplify the internal transcribed spacer (ITS) region (White et al., 1990); NL1 (5′-GCA TAT CAA TAA GCG GAG GAA AAG-3′) and NL4 (5′-GGT CCG TGT TTC AAG ACG G-3′) primers were used to amplify large subunit ribosomal (LSU) region (Kurtzman and Robnett, 1997). PCR reactions were carried out in a total mixture volume of 40 µl containing a final concentration of 20 µl DreamTag Green Master Mix (Thermo Scientific, LOT 00609058), 5 pmol of each primer, and 30-36 ng of genomic DNA. To amplify ITS region, samples were incubated in a thermal cycler (C1000 Touch, Bio-Rad) at 95°C for 3 min for initial denaturation. This was followed by 30 cycles of 95°C for 30 s, 52°C for 30 s and 72°C for 2 mins. After a final extension at 72°C for 3 min, reactions were incubated at 12°C for 5 min. To amplify LSU region, samples were incubated at 95°C for 3 min for initial denaturation. This was followed by 30 cycles of 95°C for 30 s, 57°C for 30 s and 72°C for 30s. After a final extension at 72°C for 3 min, reactions were incubated at 12°C for 5 min. The PCR products were purified using QIAquick PCR purification kit (QIAGEN, LOT157026990) following manufacturer's instructions for Sanger sequencing. Samples were sequenced using ITS1, ITS4, NL1, NL4 primers, respectively (MCLAB California, USA). Sequences were trimmed and quality checked before performing downstream analysis.

Supplemental Table 2.1. GenBank ID of LSU and ITS sequences generated from selected isolates in this study.

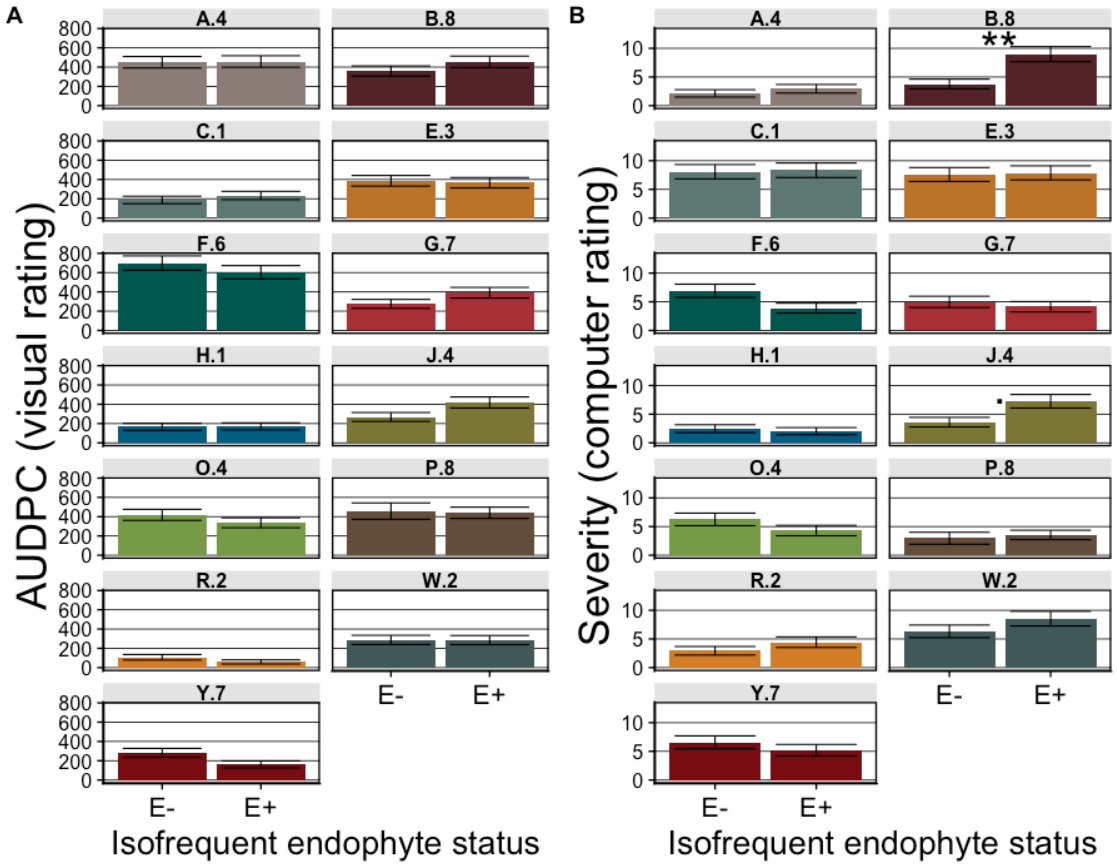
Isolate ID	LSU sequence	ITS sequence
A.1.1	MK643451	MK644123
B.1.7	MK643452	MK644124
C.1.2	MK643453	MK644125
D.1.6	MK643454	MK644126
D.1.7	MK643455	MK644127
E.1.2	MK643456	MK644128
E.1.3	MK643457	MK644129
F.1.3	MK643458	MK644130
F.1.6	MK643459	MK644131
G.1.2	MK643460	MK644132
G.1.3	MK643461	MK644133
H.1.0	MK643462	MK644134
J.1.2	MK643463	MK644135
O.1.3	MK643464	MK644136
P.1.8	MK643465	MK644137
R.1.2	MK643466	MK644138
W.1.2	MK643467	MK644139
Y.1.0	MK643469	MK644141
Y.1.7	MK643470	MK644142



Supplementary Figure 2.1 - Estimated marginal means for each of the experimental units in Experiment 1. Severity was estimated visually in the field by a single rater and regressed against image analysis scores from images taken on the same day. AUDPC calculated based on five visual scores and five image series taken over 45 days in 2017. Data shown were best modeled with a quadratic function with coefficients included in each panel.



Supplementary Figure 2.2- Plotted estimated marginal means for traits taken from five-leaf samples in Experiment 1 2017. Means were compared on an entry x genotype x endophyte basis due to a strong three-way interaction in the ANOVA. A) Total leaf area B) Percent crown rust severity based on the total area of five leaves; C) Mean crown rust pustule size; D) Mean number of crown rust pustules per leaf. Contrasts between pairs was conducted within entry (m = 5) with a Bonferroni p value adjustment: '.' = 0.1, '*' = 0.05, '**' = 0.01, '***' < 0.001. Bars surrounding means are equal to one standard error.



Supplemental Figure 2.3 - Isofrequent family pairs for Experiment 2 used to calculate narrow sense heritability. A) Plotted estimated marginal AUDPC values derived from visual scores. B) Plotted estimated marginal mean computer severity estimates in percent crown rust infection. Contrasts between pairs was conducted across entries (m = 13) with a Bonferroni P-value adjustment: '.' = 0.1, '*' = 0.05, '**' = 0.01, '***' < 0.001. Bars surrounding means are equal to one standard error.

Supplemental Figure 2.4 - Pairwise comparison of concatenated ITS and LSU sequences of the 19 isolates genotyped in this study. Overall all isolates shared a high sequence similarity. D.1.7 and A.1.1 had the lowest similarity of 99.143%. The sequence of the isolates D.1.6 and D.1.7, which were from hosts from the same entry, were 99.619 % similar. Several isolates had identical sequences.

Chapter 3

Investigating tradeoffs in perennial ryegrass turfgrass performance and seed yield capacity

Introduction

The turfgrass industry has seen tremendous gains in cultivar performance through breeding efforts focused on stress adaptation, disease resistance and quality traits (Bonos and Huff, 2013). Adoption of superior turfgrass cultivars, however, will not be successful unless the seed production capacity meets acceptable standards to both generate consistent profits for the producer while meeting consumer demand at a reasonable cost (McDonald and Copeland, 1997). Even if an increase in turf quality results in a drastic, or perhaps even a significant reduction in seed yield, seed producers may be unwilling to grow the cultivar even if there is a demand. Moreover, due to the length of time between initial spaced plant selection to final cultivar release (typically 4-8 years) any extreme trade-offs leading to unsuccessful seed production will not only reduce efficiency but also result in serious losses in revenue.

Breeding for many cool-season grasses is conducted in the northeast United States. Whereas the vast majority of seed production occurs in Oregon, Washington, and Minnesota (Bonos and Huff, 2013). It is likely that large genotype x environment interactions occur both with regard to seed production capacity as well as turfgrass quality if both parameters are observed at both locations on the same set of germplasm. For instance, orchardgrass (*Dactylis glomerata* L.) has been bred for low frequency flowering in cold short-day climates such as Wisconsin, whereas in warmer short-day climates such as Oregon higher frequencies of flowering occur (Casler et al., 2004). In this case the breeding objective was to intentionally increase seed yield plasticity between environments to maintain forage quality. Seed yield per hectare in cool-season grasses such as perennial ryegrass has dramatically increased over time through improved management of seed production fields despite selection for superior turf quality (Hart et al., 2012). However, genetic gain in either seed yield or turf quality are thought to be negatively correlated across cool-season amenity grasses with evidence available for this relationship in Kentucky bluegrass (*Poa pratensis* L) (Bonos and Huff, 2013). Johnson et al. (2003) studied the relationship between 45 populations of Kentucky bluegrass in side-by-side seed production and turfgrass environments. Results showed that seed yield had a moderate negative correlation with both color and average turf quality over a three-year span (r = -0.40 and -0.48 respectively, $P < 0.05$). The results suggest that for a single unit increase in turfgrass

quality a 21 g decrease in seed yield is expected (grand mean = 45 g) (Supplemental Fig. 3.1). However, turf quality ratings had a relatively narrow range of 4.5 to 7.2 whereas seed yields had a very wide range of 5.5 to 114 g m^{-2}. Similar results were also found in a later publication with 50 different lines of Kentucky bluegrass (Johnston et al., 2011). Żurek et al. (2007) used 27 entries of Kentucky bluegrass to measure and compare seed yield and yield components and turfgrass quality. Although no significant correlation between the two traits was found, the authors concluded that seed yield and turf quality could not be simultaneously improved. They surmised that turf affiliated traits are based on vegetative characteristics remaining consistent throughout the year, whereas seed yield depends on generative traits such as increased panicles per m^2, which could reduce quality.

No study has directly linked turfgrass performance and seed yield in perennial ryegrass. Although, Hazard et al. (2006) has been cited as providing mechanistic evidence for a potential tradeoff with turfgrass quality (Bonos and Huff, 2013). This study focused on the fitness implications of repeated defoliation of perennial ryegrass parental material by measuring progeny growth rates with respect to vegetative tillering, as well as leaf and culm elongation. Parents selected from a heavily defoliated environment produced progeny with more tillers, but displayed a slower vertical elongation rate and heading date leading to lower biomass accumulation. Progeny originating from an environment lacking defoliation elongated faster, went into a reproductive phase sooner, and were hypothesized to yield more. Unfortunately, neither seed yield nor fertile tiller number was measured making it difficult to empirically determine the impact this change in morphological development would have had in a seed production setting. That being said, it is likely that selection for slow vertical growth, maturity, and increased tillering would alter the start date and length of time spent flowering. Sampoux et al. (2013) measured seed yield and turf traits in cultivars released from 1974 to 2004 and found that no significant changes in seed yield occurred over time but increases in turfgrass quality were observed and likely were due to increases in crown rust resistance and wear tolerance. It is important to note that fungicides were applied to seed production plots in this study, which certainly would have increased seed yields if virulent pathogens were present prior to seed maturation (Pfender, 2009).

Seed production capacity is a complex trait with several yield components and subject to large genotype x environment interactions (Stratton and Ohm, 1989). Elgersma (1985) discussed seed production efficiency with respect to undomesticated grass species in general and determined that in order to increase seed yield without compromising forage or turf quality floret site utilization must be optimized. Floret site utilization in an economical sense, is the ratio of clean, harvested seeds to the number of florets per unit area. This ratio is predominantly quite low in turfgrass species; for example, in perennial ryegrass the ratio ranges from 20 to 50 % (Boelt and Studer, 2010). In theory, increasing floret site utilization would allow for a decrease in the number of spikes producing fertile tillers without compromising yield. Fertile tillers, although an important yield component of perennial ryegrass, may inherently reduce the aesthetics of the cultivar when used as a turfgrass by increasing stemmines and vertical growth rate. Sampoux et al. (2013) also noted that stemmy varieties often produced large amounts of seed but lacked turfgrass merit indicating that over emphasis on seed production may result in lack of longevity over several years.

In cold climates, winter survival of perennial ryegrass is a limiting factor in both seed production and turfgrass management. Breeding efforts that focused on increased winter survival would be a benefit for turfgrass and seed production swards as long as winter survival was not linked to poor turf quality. Hulke et al. (2007) conducted a study in which 300 accessions were tested for several turfgrass traits and found a slight negative correlation between winter survival and turf quality. Vertical growth rate and earlier progression to anthesis may benefit yield but reduce quality because of increased demand for mowing and stemminess. Increased fertile tillers is an important yield component, but increased seed head production is a negative aesthetic trait in second year turfgrass swards.

The objective of this study was to compare turfgrass quality and total seed yield of 20 perennial ryegrass entries grown in side-by-side turfgrass and seed production swards at two Minnesota locations. The experiments were designed to test several mechanistic hypotheses to determine if increasing seed yield would result in generally lower turfgrass quality. Thus, we sought to examine which in a series of eight subsidiary turfgrass traits could explain variation in the underlying relationships between turfgrass quality and seed yield.

Materials and Methods

Plant material

Perennial ryegrass experimental entries consisted of 15 interrelated populations and 5 cultivars. Interrelated populations were aggregated from several half-sib families all originating from the same source. The maternal sources for each population were selected from advanced synthetic material in the DLF Pickseed - International Seeds breeding program that was tested at the University of Minnesota and the University of Connecticut (Supplemental Table 3.1). Parental material was sent to Oregon to be polycrossed. One forage-type and four turf-type and cultivars were chosen based on a wide performance range for winter survival, turfgrass quality, and seed yield capacity (Supplemental Table 3.1).

Experimental design

Seed production and turfgrass experimental units (referred to as plots) were studied at two locations over two years. Hereafter, 2017 will be referred to as the establishment year and 2018 as the seed production year. Plots were established in early June of 2017 at both the Minnesota Agricultural Experiment Station at the University of Minnesota in St. Paul MN and the other at the University of Minnesota Magnusson Research Farm in Roseau MN. The two locations are located approximately 560 km apart and were chosen because Roseau is located within the major seed production region in Minnesota and the University of Minnesota turfgrass and seed production breeding programs are located in St. Paul. The soil at Roseau is a Borup silt loam with 2.7% organic matter and a pH of 7.7, while at St. Paul the soil is a Waukegan silt loam soil with 4.8% organic matter and a pH of 6.4.

Each of the 20 experimental entries were represented equally in both seed production and turfgrass environments. Plots for both environments were established at the same time at each location. In order to make statistical comparisons, plots of each growing environment were randomized within blocks. Experimental units were arranged as a randomized complete block within environment design with three replicates at each location.

In both years, climate data for Roseau were collected from the North Dakota Agricultural Weather Network, Fox station (48.878013, -95.847207) and the University of Minnesota St. Paul weather station (44.995031, -93.185839). Growing degree days (GDD) were calculated using a base temperature of 0 °C (Table 3.1). Soil temperature was also recorded at each location at a depth of 10 cm.

Plot design and management: Seed production environment

Seed production plots were seeded with a Hege 1000 small-plot seeder (Hege Equip. Inc; Colwich, KS). Each plot was 3.6 m long consisting of 10 drilled rows 15 cm apart. Each plot was seeded with 6.5 kg ha^{-1} of pure live seed resulting in approximately 70 seeds per linear meter of row. Each block had seeded border plots. Seed production plots were managed similarly at both locations in a manner that emulated perennial ryegrass seed production fields in Minnesota with the exception that fungicides were not applied to control disease. Upon establishment, plots were fertilized using a Gandy drop spreader (Gandy Company, Owatonna, MN) at a rate of 24.4 kg N ha^{-1}, 40.1 kg P ha^{-1}, and 24.4 kg K ha^{-1}. In late September plots were fertilized using a Gandy drop spreader at a rate of 0.0 kg N ha^{-1}, 33.6 kg P ha^{-1}, and 33.6 kg K ha^{-1}. Broadleaf weeds were controlled in the establishment year with a single application of bromoxynil and MCPA (3,5-dibromo-4-hydroxybenzonitrile and 2-ethylhexyl ester of 2-methyl-chlorophenoxyacetic acid; Bison, BASF, Research Park Triangle, NC) at a rate of 0.28 kg a.i. ha^{-1} at each environment. Fall and early spring seedlings were controlled with a single application of pendimethalin (N-(1-ethylpropyl)-3,4-dimethyl-2,6-dinitrobenzenamine; Prowl H_2O, BASF, Research Park Triangle, NC) at a rate of 2.12 kg a.i. ha^{-1} at each environment. Fall broadleaf weeds were controlled with an application of 2,4-D (2, 4-dichlorophenoxyacetic acid; Shredder, Winfield St Paul, MN) at a rate of 0.40 kg a.i. ha^{-1}. In the seed production year fertilizer was applied in early May at a rate of 123.1 kg N ha^{-1}. Broadleaf weeds were controlled in late May with an application of 2,4-D (2, 4-dichlorophenoxyacetic acid; Shredder, Winfield St Paul, MN) at a rate of 0.27 kg a.i. ha^{-1} and dicamba (3,6-dichloro-o-anisic acid; Sterling Blue, Winfield St Paul, MN) at a rate of 0.27 kg a.i. ha^{-1} at each environment. All chemical applications were applied using a CO_2-powered bicycle sprayer equipped with a 2.75 m boom and eight 1002 TurboTeeJet® (TeeJet, Springfield, IL)

nozzles operating at 186 kPa in water equivalent to 117 L ha^{-1}. Sampling for seed yield occurred in early July in St. Paul and early August in Roseau. Sampling was conducted twice at St. Paul and Roseau to ensure that ripe seed at approximately 400 g kg^{-1} moisture content was harvested. Yield samples entailed harvesting two 1 m^2 swaths per experiential unit at approximately 400 g kg^{-1} moisture content. Biomass was dried for 5 d at 35 °C in a forced air dryer.

Plot design and management: turfgrass environment

Turfgrass plots were 0.91 x 0.91 m. Seeding was done by hand at a rate of 90 kg ha^{-1} pure live seed resulting in approximately 0.5 seeds cm^{-2}. Turfgrass plots received the same establishing fertilizer and chemical regime as the seed production plots. Turfgrass plots were fertilized in May 2018 using a Gandy drop spreader (Gandy Company, Owatonna, MN) at a rate of 24.4 kg N ha^{-1}, 40.1 kg P ha^{-1}, and 24.4 kg K ha^{-1} at both locations. Plots at Roseau received an additional application at the same rate in July in order to facilitate regrowth from severe winterkill and a lower soil organic matter. Turfgrass was maintained at a height of 6.7 cm by mowing every other week during the growing season.

Data collection: Seed production environment

Response variables for the seed production swards in the establishment year included stand counts and image data. Seedling emergence was estimated three weeks after seeding by counting the number of seeds that emerged in two 0.5 m transects. Image data were collected every two weeks at both locations three weeks after seeding to measure crown rust incidence (image analysis technique described below).

Response variables in the seed production year included winter survival, plant height, fertile tillers m^{-2}, and seed yield m^{-2}. Winter survival was quantified using digital image analysis, with the proportion living tissue to dead tissue within row equal to survival. Plant height was measured after anthesis by selecting culms at three random points in each plot. Fertile tillers m^{-2} was estimated by cutting four 15 cm transects of culms in each plot. Each of these transects were cut twice so that 0.25 cm pieces of culm were isolated from the bottom of each transect and funneled into a coin envelope. Culm pieces were cleaned

with a general seed blower to remove leaves and other debris and counted using a Seed Count S-JR seed counter (Wintersteiger Corp., Salt Lake City, UT). Weight of cleaned seed was measured after processing each sample in a LD350 thresher (Wintersteiger Corp., Salt Lake City, UT). Seed and chaff were separated by passing the seed through a 1.625 x 9.525 mm sieve. Light seed was removed by aspiration using a Superior® Fractionating Aspirator (Carter-Day Int. Inc., Minneapolis, MN).

Data collection: turfgrass environment

Turfgrass sward establishment was assessed on a 1-9 visual scale with a 5 equal to 50% of the seedlings filling in a 15 x 15 cm square grid. Turfgrass quality, genetic color, texture, density, rust incidence, height, and image data were collected both years. All turfgrass quality traits were assessed according to the National Turfgrass Evaluation Program guidelines (Morris and Shearman, 2006). Custom diagrammatic scales were used for density and genetic color to maintain consistent rating across locations (Supplemental Fig. 3.2). Turf height was assessed before each mowing by measuring three points in each plot using a TurfChek II grass height gauge (Turf-Tec International, Tallahassee, FL). Images data was collected every two weeks at both locations three weeks after seeding to measure growth rate, crown rust infection, and winter survival (see image analysis technique). Stemminess can be measured visually (Islam et al., 2013), but was also quantified through visual stem counts and the proportion of stems detected in images.

Image analysis technique

Images were taken with a Nikon D300 digital SLR camera and saved as high quality JPEG files of 4288 x 2848 pixels. Imaging was conducted under ambient conditions in the establishment year between 1200 and 1400 hours with a fixed monopod mounted 80 cm above the ground for turf plots and 160 cm for seed plots, squarely above the plot. Camera settings were shutter speed of 1/160 s, aperture setting of F8, white balance 5000K, and a focal length of 80 mm. In 2018, images were taken using the same camera but with a light box constructed of an aluminum frame (0.8 m wide x 0.6 m long x 0.9 m tall) with coroplast walls lined with mylar to eliminate light infiltration. The interior was lit by 200 cm of high color rendering index LED light strip with an output of 221 lumens 0.3 m^{-1} and a 4000K

color value (superbrighLED.com). The lights were powered by a rechargeable Talentcell 12-volt lithium ion battery pack (talentcell.com). Camera settings for light box images were shutter speed 1/80 s, aperture setting 2.7, white balance 5000K, and a focal length of 80 mm.

Quantification of green cover was conducted using an automated system described by Heineck et al. (2019) for both turfgrass and seed environments. Image analysis was done using ImageJ and R (Schneider et al., 2012; Version 3.6.0; R Core Team, 2018). Core packages in R were EBImage and randomForest (Liaw and Wiener, 2002; Pau et al., 2010). Separate training data sets were made for each year and environment using images from each rating date with about 300 data points were collected from each image to train the model. Individual random forest models were generated for each year and environment. Image analysis output at each time point was inspected visually for any abnormalities in the processing pipeline. The proportion of green pixels was collected as a repeated measure.

Statistics

Analysis of all data was conducted in the R environment (R Core Team, 2018). Analysis of variance and analysis of covariance was conducted using a fixed effect model in the 'agricolae' package (Mendiburu, 2016). Turfgrass quality was modeled as a repeated measure with time treated as a continuous variable. Marginal means were estimated, and separations were conducted using the 'emmeans' package and a Tukey HSD test to detect significant differences ($\alpha = 0.05$) (Lenth, 2018). Bivariate relationships were explored through simple linear regression with large numbers of regressions conducted using the 'PerformanceAnalytics' package. Correlation coefficients were based on Pearson product moment calculations.

The relationship between turfgrass quality and seed yield was measured both on a grand mean and by individual rating dates. When regression was conducted between turfgrass and seed environments entry means were used; however, when regression was done within environments data points from each block were used. Influence on seed yield was determined though linear regression with moderate significance indicating need for further investigation ($\alpha = 0.1$). Turfgrass quality rating dates of interest were explored

through multiple regression with subsidiary traits included as explanatory variables and scaled by subtracting the mean and dividing by the standard deviation of each variable. Optimum model selection for multiple regression was conducted with the 'leaps' package set perform an exhaustive search with adjusted R^2 determining the optimal combination of explanatory variables (Lumley and Lumley, 2013). The importance of each scaled variable in the optimal model was measured through multiple regression and partial correlation coefficients ($\alpha = 0.05$).

Results

Growth during the establishment year

Turfgrass plots established successfully at both locations. No significant differences for turfgrass plot establishment were found among entries at or between Roseau or St. Paul. Entries differed in both vertical and lateral growth rates at Roseau ($P < 0.001$) and at St. Paul ($P < 0.001$). Lateral growth in turfgrass plots during establishing year resulted in 80% green coverage at Roseau and 75% at St. Paul averaged

Seed production plots at St. Paul obtained a higher plant population than Roseau (49 vs. 32 plants m row^{-1} respectively); however, no differences were found between entries within location. Significant differences among entries for biomass accumulation were found in seed production plots at St. Paul ($P < 0.001$), but not Roseau ($P = 0.494$). Chlorophyll index measures on an entry mean basis were significantly different at both St. Paul ($P < 0.001$) and at Roseau ($P = 0.04$). Crown rust was present at both locations, and entries differed significantly in severity estimates in both turfgrass and seed production plots ($P < 0.001$).

Winter survival

Winter survival was determined through digital image analysis in seed production and turfgrass plots. Winter conditions were much more extreme at Roseau compared to St. Paul with far more subzero air and soil temperatures and colder extreme minimum temperatures (Table 3.1). Although St. Paul experienced some winter injury, tiller mortality was much greater in Roseau both in the turfgrass and seed production environments.

Winter survival did not differ significantly among entries in the turfgrass environment at either location. Turfgrass plots averaged 31% survival in Roseau and 52% at St. Paul.

Entries in seed production environments differed significantly for winter survival at both locations. Survival on an entry mean basis at Roseau ranged from 1 to 16 % and 29 to 74 % at St. Paul. It is worth noting that seed production environments have over seven times the plot area of turfgrass environments and therefore would have been less susceptible to random spatial winterkill effects. Winter survival on an entry mean basis between turfgrass and seed production plots were positively correlated at Roseau ($r = 0.44$, $P = 0.05$) and at St. Paul ($r = 0.47$, $P = 0.01$). However, these models both explained less than 30% of variance.

Turfgrass quality

Turfgrass quality was measured four times in the establishment year and five times in the seed production year. Analysis of covariance showed that although large differences were found among entries in general, there were strong interactions between years and rating dates (Table 3.2). Furthermore, rating date was best modeled with a quadratic term which interacted with different entries at different times depending on the year. At Roseau, turf quality generally increased over time in both years, however quality scores were much lower in the production year (Fig. 3.1). At St. Paul, turfgrass quality scores in the establishing year increased until mid-August and then decreased late in the season. Average turf quality at St. Paul remained similar in the production year and were stable throughout the growing season (Fig. 3.1).

Although entries did not differ significantly for winter kill, tiller survival was related to turfgrass quality on a plot basis at both locations (Fig. 3.2A and B). Turfgrass quality was best modeled using a quadratic model at both locations with survival predicting 84 and 32% of variance for turfgrass quality in Roseau and St. Paul, respectively. As winter survival decreased turfgrass quality increased to a certain extent and then plateaued off around 70% survival. This was much more evident at St. Paul simply because Roseau plots never exceeded 60% survival.

Seed yield

Differences in seed yield were found among entries at both locations, with far lower variation within entry at St. Paul (CV = 10.6%) compared to Roseau (CV = 49.0%) (Supplemental Figure 3.3). Entry explained the majority of variance for yield, but winter survival played a significant role as well (Table 3.3). Analysis of covariance found that winter survival at St. Paul, although significant, accounted for only 8% of variance for seed yield in the model, whereas entry accounted for 65%. Winter kill was much more pronounced at Roseau with winter survival accounting for 28% of variance for seed yield, whereas entry accounted for 48 %. Winter survival was similar to turfgrass quality as it too had a quadratic relationship with seed yield. Polynomial regression showed that as tiller survival approached and exceeded 20%, seed yield was only moderately unaffected (Fig. 3.2C and D). Few plots at St. Paul were below 20 % survival and the slope of the model was positive but not significant ($P = 0.32$).

Fertile tiller number was much higher in St. Paul compared to Roseau (6400 vs. 1400 spikes m^{-2} respectively). There was no significant effect of entry on fertile tiller number (FTN) in St. Paul ($P = 0.41$), but entries did differ somewhat for FTN in Roseau ($P = 0.04$). Seed yield was positively correlated with FTN at both locations on a per plot basis ($P < 0.006$). Winter survival and FTN were highly correlated at Roseau ($r = 0.85$, $P < 0.001$), but not at St. Paul. Growth stage was measured multiple times to assess differences in culm elongation and flowering time. Larger growth staging values indicate earlier flowering. There were significant differences among entries at both locations for both growth stage ratings ($P < 0.05$). At Roseau, seed yield was positively correlated growth stage at both rating dates ($P = 0.03$ and 0.001, respectively). At St. Paul, growth stage was only correlated with yield at the second rating date ($P = 0.96$ and 0.001).

Relationship between turf quality and seed yield

There was no significant relationship between seed yield and overall average turf quality at either location (Fig. 3.3). Turfgrass quality ratings through time interacted with both year and entry (Table 3.2). Therefore, each rating date in the establishment and seed production year was regressed against seed yield systematically (Supplemental Fig. 3.4). At Roseau, the third rating in the establishing year and the first and second in the production year were moderately correlated with seed yield ($r = 0.41$, 0.41, and 0.39 $P < 0.1$,

respectively). At St. Paul, the fourth rating in the establishment year and fifth in the production year were moderately correlated with seed yield (r = 0.42 and 0.40, $P < 0.1$, respectively).

These five rating dates were explored to find subsidiary turfgrass quality traits of importance that could explain potential mechanistic reasons behind the significant correlation with seed yield. Each of these correlation coefficients was positive, and it is interesting to note that there was not a single negative coefficient at any rating date. Turf quality related traits in the establishment year included: establishment, vertical growth, lateral growth, genetic color, texture, chlorophyll index, and crown rust severity. Turfgrass quality traits in the production year included: winter survival, vertical growth, lateral growth, genetic color, chlorophyll index, and proportion of dried stems (steminess). Each of these traits was exhaustively added to a multiple regression model that determined the most important variables on a plot basis based off of adjusted R^2.

In the establishment year, turf quality at Roseau was most influenced by establishment (σ = -0.23, P = 0.005), genetic color (σ = 0.28, P = 0.0007), crown rust severity (σ = -0.27, P = 0.007), and lateral growth (σ = 0.55, $P < 0.001$). In the production year, quality was most impacted by winter survival was most important at the first and second rating dates (σ = 0.83 and 0.61, $P < 0.001$, respectively). However, in the second rating date chlorophyll index (σ = 0.29, P = 0.002) and proportion of stems (σ = -0.23, $P <$ 0.001) were found to be related to turf quality.

In the establishment year, turf quality at St. Paul was the most associated with crown rust severity (σ = -0.35, $P < 0.001$), followed by lateral growth (σ = 0.46, $P < 0.002$), and chlorophyll index (σ = 0.30, $P < 0.02$). In the production year at St. Paul, vertical growth, lateral growth, chlorophyll index, and proportion of stems were important. However, none of these explained a substantial amount of variation with chlorophyll index being the only significant predictor (σ = 0.39, P = 0.04). Vertical growth (σ = -0.21, P = 0.09) and proportion of stems (σ = -0.17, P = 0.25) both had a marginally negatively impact on turf quality.

Relationship between fertile tiller number and stemminess

Fertile tiller number was estimated in the seed production year in both the seed production and turfgrass environments. Fertile tiller number in seed production plots was positively correlated with yield on a per plot basis at both environments (Fig. 3.4A). The relationship at Roseau was highly significant indicating that fertile tillers were limiting for seed yield in this environment. A negative correlation between proportion of stems in turf plots and quality scores was present at both locations on a per plot basis (Fig. 3.4B). Although fertile tillers positively correlate with yield and negatively correlate with turf quality the proportion of stems in turfgrass plots was not correlated with fertile tiller number in seed plots at either location (Fig. 3.4C).

Relationship between growth rate and maturity

In the establishment year, turfgrass lateral growth was strongly correlated with biomass accumulation in the seed production environment at both Roseau ($r = 0.51$, $P = 0.02$) and St. Paul ($r = 0.82$, $P < 0.001$). Vertical turfgrass growth was not correlated with biomass accumulation in seed plots at either location (Supplemental Fig. 3.5). Correlations for crown rust severity between turfgrass and seed production environments were extremely significant at both Roseau ($r = 0.95$, $P < 0.001$) and at St. Paul ($r = 0.99$, $P < 0.001$).

Turfgrass lateral growth was negatively correlated with vertical growth at both locations in the seed production year (Supplementary Fig. 3.5). Vertical growth rate was strongly correlated with more advanced growth stages in the seed production environment at St. Paul ($r =0.56$, $P = 0.01$), but not at Roseau. In seed production plots earlier flowering (more advanced growth stage) was associated with increased seed yield. However, despite this faster vertical turf growth being associated with earlier flowering, growth rate in turfgrass plots was not associated with higher seed yields ($r = 0.05$, $P = 0.82$).

Discussion

Although seed yield capacity is widely recognized as an important trait, methodical observation and recording of gains in breeding programs has not been well documented in the turfgrass industry (Casler et al., 1996). Management practices have been well researched and have led to marked increases in production capacity both in Oregon and

Minnesota (Hart et al., 2012; Koeritz et al., 2015). Often screening of new germplasm for seed yield capacity does not occur until late in the breeding program, this can lead to serious consequences if the cultivar does meet economic yield demands. Breeding for both elite turfgrass quality and seed yield can be especially difficult in Kentucky bluegrass, but this dogma extends to other species such as perennial ryegrass (Meyer and Funk, 1989; Bonos and Huff, 2013). This study sought to explore the relationship between turfgrass and seed yield traits in Minnesota using both interrelated germplasm and common check cultivars. The two study locations in this experiment experienced vastly different growing conditions with the more northern location having frozen soil for 155 d and very little rainfall in the spring of the seed production year. These conditions led to massive winter kill the need to analyze and interpret results from each location individually. There was significant variation among entries for average turf quality scores and seed yield (Tables 3.2 and 3). However, there was no direct relationship between these two traits at either location (Fig. 3.3). Turfgrass quality was observed as a repeated measure and is also a complex trait composed of at least six sub traits (Morris and Shearman, 2006). Our analysis began with exploring which rating dates were most related to seed yield and then determine if trends existed that would indicate the importance of subsidiary traits. These traits (e.g. genetic color, vertical growth rate, disease severity) could then be used to test potential tradeoffs between the improvement of certain aspects of turfgrass quality and seed yield that may have been masked by the grand mean. For instance, do important generative traits such as timing and speed of stem elongation or fertile tillering related positively to seed yield, but negatively to turf quality (Żurek et al., 2007)?

Turfgrass quality ratings for each entry were highly variable over years and environments, with different stressors presumably impacting quality across different rating dates. There was a significant interaction between years, entry and rating time; however, there was not a significant interaction between entry and rating time within year (Table 3.2). The general interpretation of this may be that entries followed similar trends each year, but trends changed between years and time points within years. In 2017 turf quality at Roseau steadily increased throughout the growing season, whereas quality at St. Paul began to decline late in the growing season (Fig. 3.1). Turfgrass quality at Roseau was very low in 2018 compared to 2017. In 2018, turf quality at St. Paul remained fairly consistent

with several entries maintain an acceptable quality score (>6.0). Because of the strong interaction between rating date and year it was important to both test individual rating dates with respect to their relationship with seed yield. Furthermore, isolating individual traits that significantly impact turf quality allowed for a more complete understanding of potential tradeoffs. This assumption is in contrast to Johnson (2006) who observed the mean turf quality rating of plots over three years and found a seemingly simple negative relationship between quality and yield.

When each turfgrass rating date was correlated against seed yield, several dates showed only moderate positive significance (r = 0.4 to 0.42, $P < 0.1$), but important rating dates were not always consistent between years and locations. For instance, the latter ratings of the establishment year were important at both locations, however the initial ratings in the production year were important at Roseau and only the last at St. Paul was correlated with yield. To further explore commonalities between dates of interest, exhaustive multiple regression analysis was used to determine which principle turfgrass traits were important at each date. At both locations in the establishment year, increased turf quality of the later ratings dates was driven by crown rust severity (negative) and lateral growth (positive). Entries significantly differed for crown rust severity in both turfgrass and seed production growing environments at both locations ($P < 0.001$). Furthermore, severity estimates between environments were highly correlated on an entry mean basis at Roseau (r = 0.96, $P < 0.001$) and St. Paul (r = 0.98, $P < 0.001$). These near perfect correlations give strong evidence that if selection pressure is imposed for these traits in a turfgrass environment the same result should be mirrored in the seed production environment. Correlation between growing environments was also found between lateral growth and biomass accumulation in turf and seed production plots respectively at Roseau (r = 0.51, $P < 0.02$) and St. Paul (r = 0.82, $P < 0.001$). Although these may seem like obvious relationships, this does clearly demonstrate that improvement of several key turf traits would have a favorable influence in a seed production environment.

In the seed production year, the first two rating quality dates were associated with seed yield at Roseau, where increased winter survival was positively correlated with turfgrass quality. This is not surprising as both average turfgrass quality and seed yield were highly influenced by winter kill at Roseau (Fig. 3.2). This was not the case in St. Paul

where only turfgrass quality was correlated to winter kill (Fig. 3.2). It is interesting to note that average turfgrass quality was altered when tiller mortality was under 60 %, whereas seed yield did not respond until tiller survival was under 20 %. This difference in response to winter damage was likely due seed plots increasing fertile tillers to compensate for damage by the end of the season. This data shows that winter survival is important to both maximize seed yield and turf quality in the Midwest. This also demonstrates that improving germplasm winter hardiness will not result in a negative tradeoff. Furthermore, no consistent associations between tiller survival and turfgrass quality or other traits such as genetic were found. Hulke et al. (2007) found that increased winter survival was negatively correlated with turfgrass quality and genetic color on a spaced plant basis. This was certainly not the case here; however, sward environments and only elite material was tested in this study.

At St. Paul, in the seed production year the final rating date correlated with seed yield. Vertical growth (negative) and chlorophyll index (positive) were associated with turfgrass quality. Vertical growth also negatively influenced turf quality in Roseau. These results gave evidence that reducing vertical growth rate and stemminess may have some minor tradeoff with seed yield. Early flowering was correlated with increased seed yields at both locations. This trait has been generally associated with increased seed yields in the past (Elgersma et al., 1994). At St. Paul, growth stage and vertical growth rate in turfgrass plots were positively correlated; however, this trend was not observed at Roseau. Vertical growth rate and seed yield however had no direct relationship (Fig. 3.4) leading to the conclusion that reducing vertical growth rate will likely increase time to flowering thereby possibly negatively effecting seed yield.

Fertile tiller number is an important seed yield component and has been related to seed yield under certain situations (Bugge, 1987; Chastain and Young III, 1998). Perennial ryegrass has extremely low juvenility with fall produced primary tillers able to initiate a phase change from vegetative to reproductive to secondary spring produced tillers (Onishi et al., 2003). This mechanism likely allows perennial ryegrass to recover from substantial winter damage with little to no yield reduction (Fig. 3.2). However, in a turf environment fertile tillers quickly become visible post mowing leaving a straw-colored stubble. Once dry, these stems reduce the aesthetics of a perennial ryegrass stand in the second year.

Therefore, a direct tradeoff between seed yield potential and turfgrass quality may be realized through breeding for decreased fertile tillering. In the seed production year, when turfgrass quality was associated with seed yield, proportion of stems in turfgrass had a negative influence on turfgrass quality at that rating date. On a per plot basis fertile tiller number was highly correlated with seed yield at both locations. In turfgrass plots, stemminess was negatively correlated to turfgrass quality (Fig. 3.3). Because of the positive correlation between culms and seed yield and the negative correlation between proportion of stems and turf quality is seemed likely that turf plots with lots of stems would result in seed plots of the same entry having large numbers of fertile tillers. Surprisingly, fertile tiller number and proportion of stems in a plot were not correlated with each other on an entry mean basis (Fig. 3.3C). This would suggest that breeding for decreased for stemness would not result in a change in fertile tiller number in a plot environment.

It is important to remember that there was not a single instance in which turf quality negatively correlated with seed yield. Certain traits may have been associated with tradeoffs, but the negative partial correlations were never prominent enough to make and overall negative association. There may be a tradeoff when breeding for reduced vertical growth in the seed production year, however these results were inconsistent. In relation fertile tillering and steminess, it could be that winter damage increased the importance of spring tillering leading to more fertile tillers and yield. This may be confounding the relationship between FTN and stemminess. It should be noted that this study is somewhat limited by both the number of entries included and the population design. Although 15 interrelated populations were used along with common check cultivar a divergent selection scheme should be conducted on a population using bulk recurrent selection using progeny tests for both turfgrass quality and seed yield.

Conclusion

This study explored potential tradeoffs between turfgrass quality and seed production capacity in perennial ryegrass. Turfgrass quality was regressed against seed yield both on a grand mean and on an individual rating date scale. Overall no correlation existed, however, several individual rating dates were moderately correlated. Several traits such as crown rust, winter survival, proportion of dried stems, lateral and vertical growth,

were found to consistently explain variance in models including rating dates of interest. Both crown rust and winter survival were positively correlated between turfgrass and seed production growing environments. Winter survival negatively influenced both quality and yield similarly and may partially explain the slight positive trend between yield and turf quality. Hypotheses including the implications of reducing vertical elongation and stemminess were then explored. Fertile tillers were not associated with greater proportions or counts of stems in turf plots. Vertical elongation in turf plots was highly correlated with rapid growth stage advancement in seed plots. Although this was not directly impactful on yield it could be indirectly influential. In northern climates it is unlikely that any major tradeoff between turf quality and seed yield exists.

Tables

Table 3.1 – Summary of growing conditions at St. Paul and Roseau during the establishing year (2017) and seed production year (2018). Growing degree days were calculated with a base temperature of 0 ºC.

	Summer 2017†		Winter 2017-18‡			Summer 2018¶	
	GDD	Precip (cm)	Min Soil °C	Soil d < 0 °C	Air d < 0 °C	GDD	Precip (cm)
Roseau	2043	33.8	-4.7	155	164	1973	18.2
St. Paul	2406	40.0	-2.2	110	134	2487	36.8

† Summer 2017: June 1st - September 30th

‡ Winter 2017-18: October 1st - April 15th

¶ Summer 2018: April 16th - August 15th

Table 3.2 - Analysis of covariance for turfgrass quality scores taken as a repeated measured across years within each location. Date (D) was nested within. Date was modeled as a continuous variable that was best fit with a quadratic term.

		St. Paul			Roseau		
Source	Df	MS	F-value	P-value	MS	F-value	P-value
block	2	25.8	8.8	<0.001	17.9	4.2	0.021
entry (E)	19	19.1	6.5	<0.001	10.5	2.5	0.008
Error A	38	2.9	-	-	4.2	-	-
year (Y)	1	3.6	6.8	0.009	1114.3	1825.6	<0.001
E:Y	19	3.2	6.1	<0.001	2.6	4.3	<0.001
Y:D:D^2	1	3.9	7.4	0.002	35.3	57.8	<0.001
E:D:D^2	19	0.3	0.7	0.732	1.1	1.8	0.949
E:Y:D:D^2	19	1.1	2.2	0.001	2.2	3.5	<0.001
Error B	421	0.5	-	-	0.6	-	-

Table 3.3 - Analysis of covariance seed yield at each location. Winter survival was included as a covariate due to its significance in each model.

		St. Paul			Roseau		
Source	Df	MS	F value	P value	MS	F value	P value
block	2	1018.81	9.5	0.001	261.1	9.4	0.002
entry (E)	19	655.2	6.1	<0.001	221.2	7.9	<0.001
green-up (G)	1	968.9	9.1	0.007	2432.7	87.4	<0.001
E:G	19	96.6	.9	0.583	53.8	1.9	0.084
Error	18	106.4	-	-	27.8	-	-

Figures

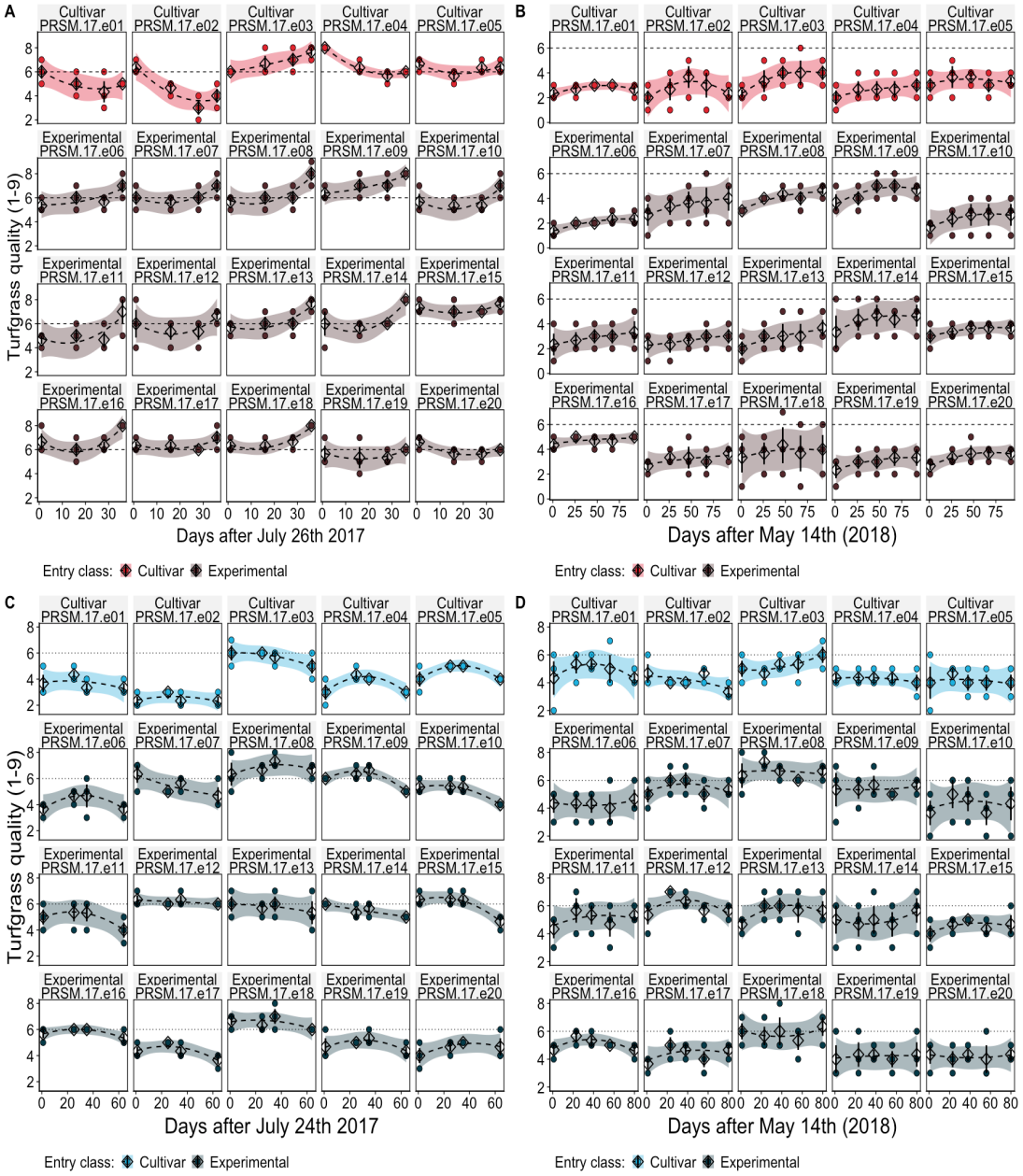


Figure 3.1 - Turfgrass performance for each year and location. Rating began on 7/26/17, 5/31/18, 7/24/17, and 5/14/18 for A, B, C, and D respectively. Dots are the observed value for each plot, triangles with bar are the estimated marginal mean +/- 1 standard error, dashed line with shaded area is the predicted quadratic curve +/- 1 standard error. The dotted black line is a turfgrass quality score of 6, which is equal to acceptable quality.

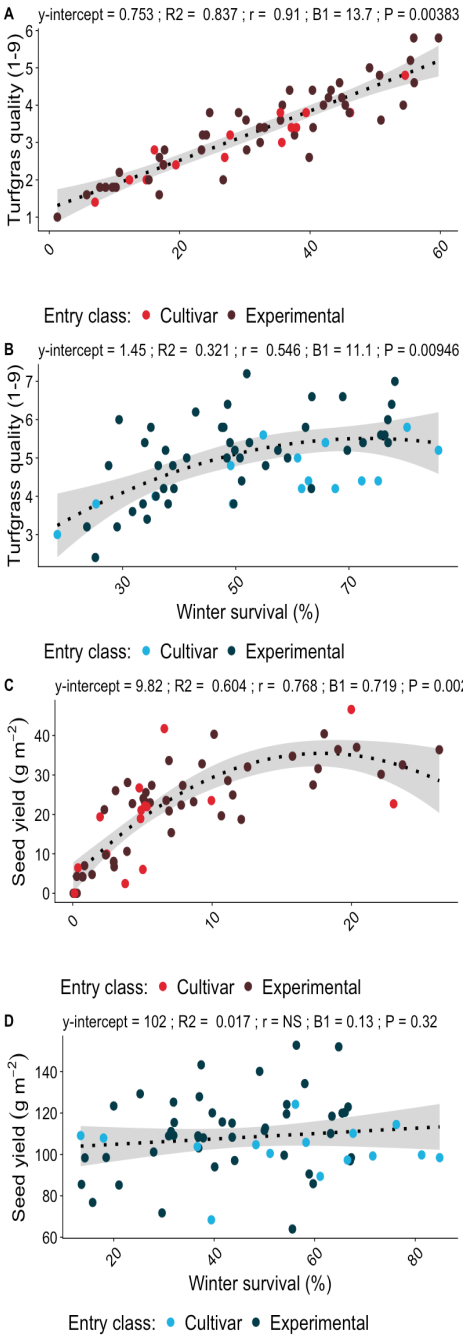


Figure 3.2 - Linear and polynomial regression for winter survival against turfgrass quality and seed yield. A) linear relationship between survival and turf quality at Roseau. B) Polynomial relationship between survival and turf quality at St. Paul. C) Polynomial relationship between survival and seed yield at Roseau. D) Insignificant relationship between survival and sed yield at St. Paul.

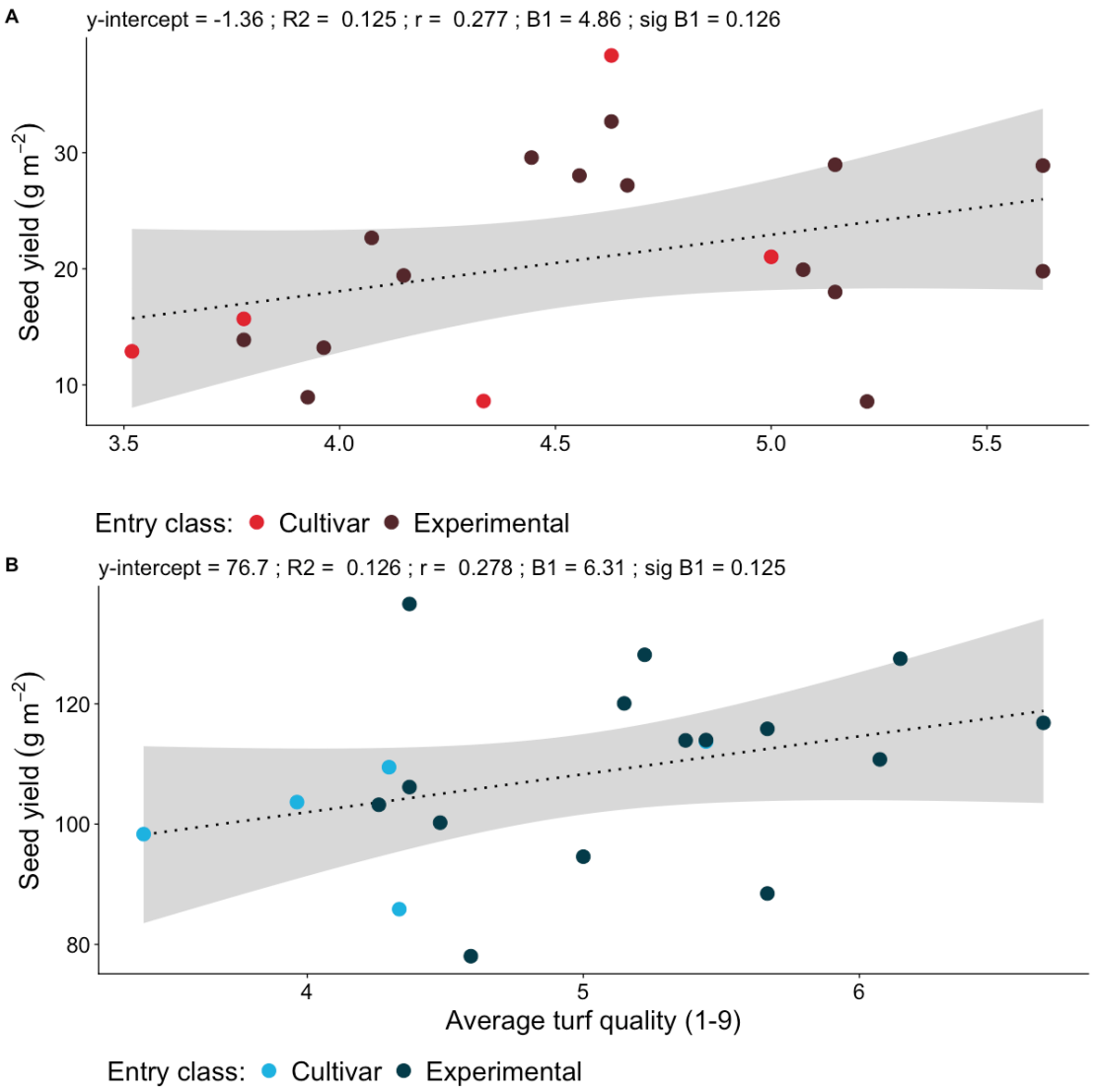


Figure 3.3 - Relationship between turfgrass quality and seed yield at Roseau (A) and St. Paul (B). Turfgrass quality on an entry mean basis was averaged across rating dates within location. Dotted line represents the linear prediction with the shaded region equal to +/- standard error.

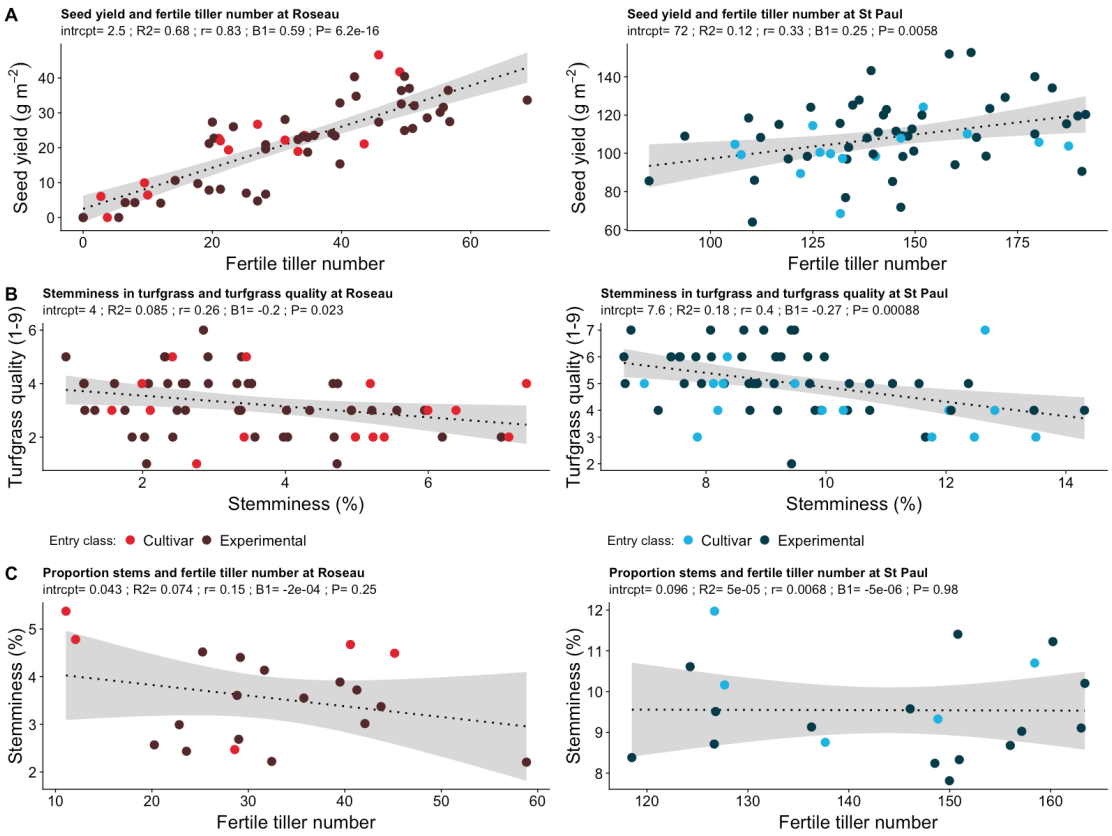


Figure 3.4 - Relationship of fertile tiller prevalence on turf quality and seed yield. Fertile tiller number is a seed yield component in seed production plots. Proportion of stems is an unfavorable trait in turfgrass environments. A) Positive linear relationship between fertile tiller number and seed yield. B) Negative relationship between turfgrass quality and proportion of stems. C) Insignificant relationship between fertile tiller number in seed plots and proportion of stems in a turfgrass plots.

Supplemental

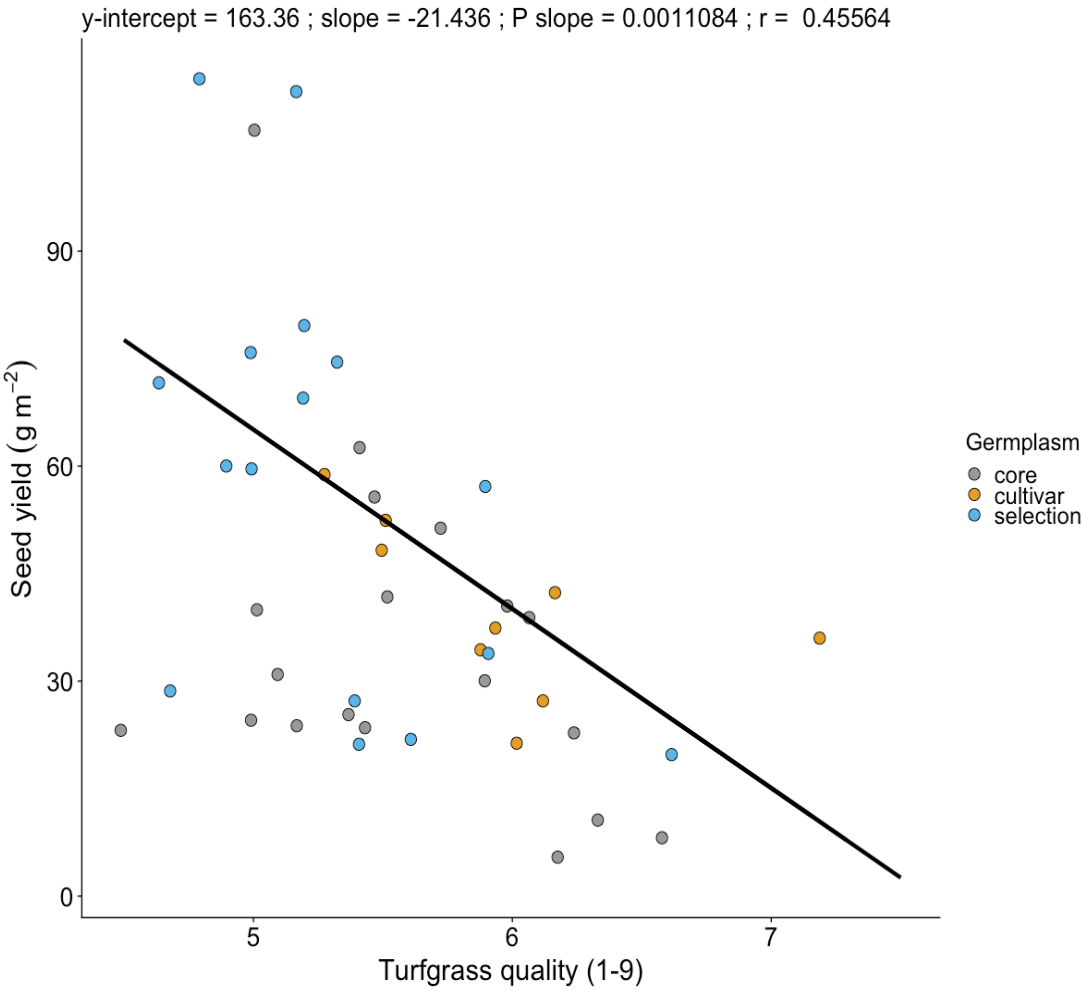
Supplemental table 3.1- Experimental entries included in both the turfgrass and seed production trials. Check entries were commercial cultivars and include information on turfgrass quality, seed yield, and winter hardiness. Experimental entries were interrelated half-sib populations donated by DLF International Seeds breeding program with parental sources in commercial cultivars and advanced breeding populations.

Entry designation	Cultivar/selection name	Germplasm type	Seed source	Parental source	Turfgrass quality rating †	Relative seed yield 2 ‡	Winter hardiness §
PRSM.17.e01	Quebec	Check	DLF		3.1	NA	-2.7
PRSM.17.e02	NK 200	Check	University of MN		NA	Poor	0
PRSM.17.e03	Metolius	Check	Peak Plant Genetics		6.1	NA	NA
PRSM.17.e04	Fiesta_4	Check	DLF		4.7	Moderate	-2.4
PRSM.17.e05	Arctic_Green	Check	University of MN		4.3	Excellent	-0.7
PRSM.17.e06	NGSPRWH-2-12	Experimental	NA	NA	NA	NA	NA
PRSM.17.e07	PR-24-16-06	Experimental	Connecticut	PSG.1037-12K	NA	NA	NA
PRSM.17.e08	PR-24-16-08	Experimental	Connecticut	Zoom	NA	NA	NA
PRSM.17.e09	PR-24-16-09	Experimental	Minnesota	PR.580	NA	NA	NA
PRSM.17.e10	PR-24-16-12	Experimental	Connecticut	Bandalore	NA	NA	NA
PRSM.17.e11	PR-24-16-14	Experimental	Connecticut	PR.562	NA	NA	NA
PRSM.17.e12	PR-24-16-17	Experimental	Minnesota	Karma	NA	NA	NA
PRSM.17.e13	PR-24-16-21	Experimental	Minnesota	PR.524C	NA	NA	NA
PRSM.17.e14	PR-24-16-22	Experimental	Connecticut	Diligent	NA	NA	NA
PRSM.17.e15	PR-24-16-23	Experimental	Minnesota	Monsieur	NA	NA	NA
PRSM.17.e16	PR-24-16-25	Experimental	Minnesota	Aspire	NA	NA	NA
PRSM.17.e17	PR-24-16-26	Experimental	Connecticut	PSG.HLT	NA	NA	NA
PRSM.17.e18	PR-24-16-28	Experimental	Connecticut	PR.583	NA	NA	NA
PRSM.17.e19	PRWHNGS-1-14	Experimental	NA	NA	NA	NA	NA
PRSM.17.e20	WH-PR-11-03	Experimental	NA	NA	NA	NA	NA

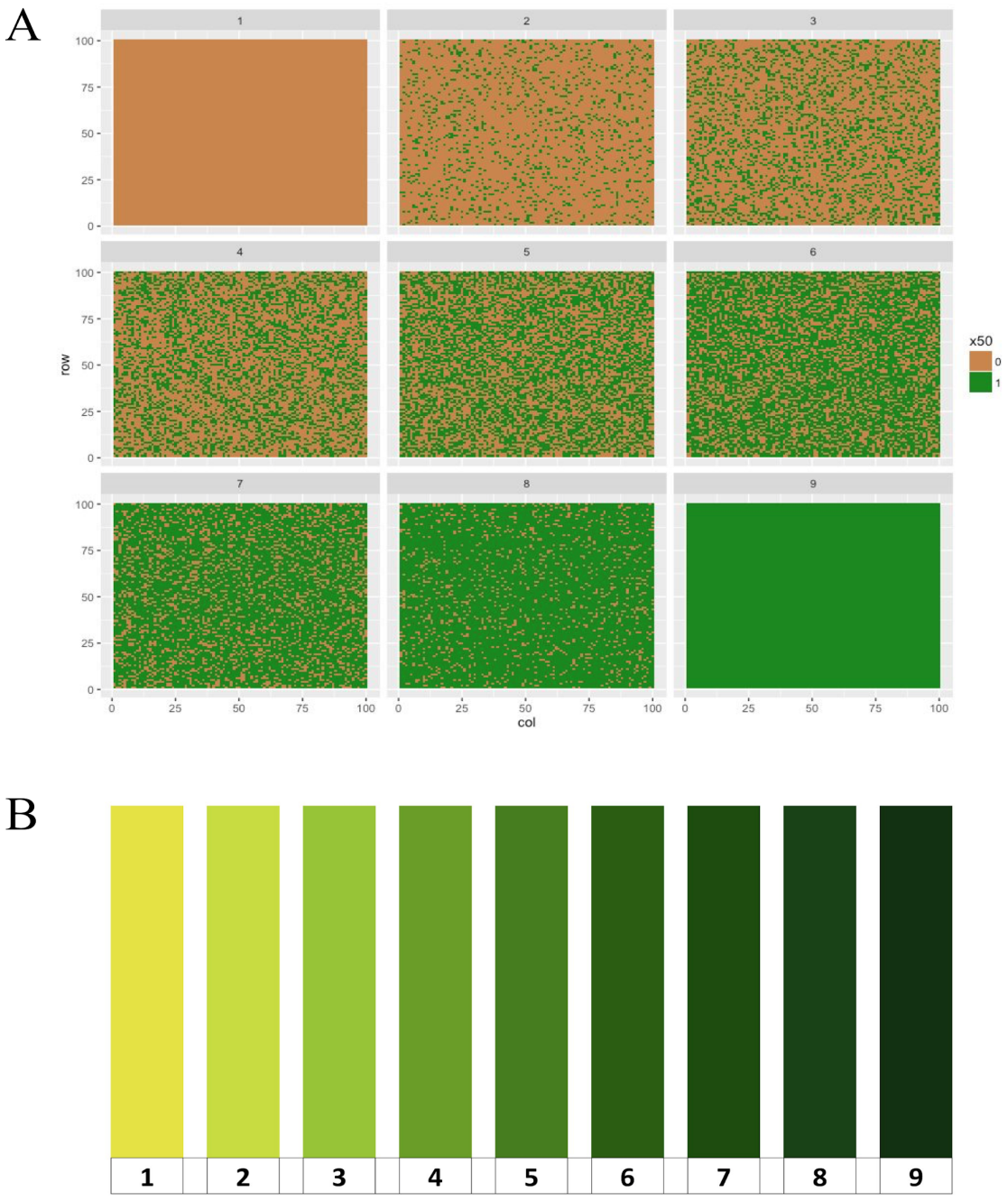
† Turfgrass quality determined by national turfgrass trial data.

‡ Seed production capacity determined by MTSC reports relative to NK200 (2013-2017).

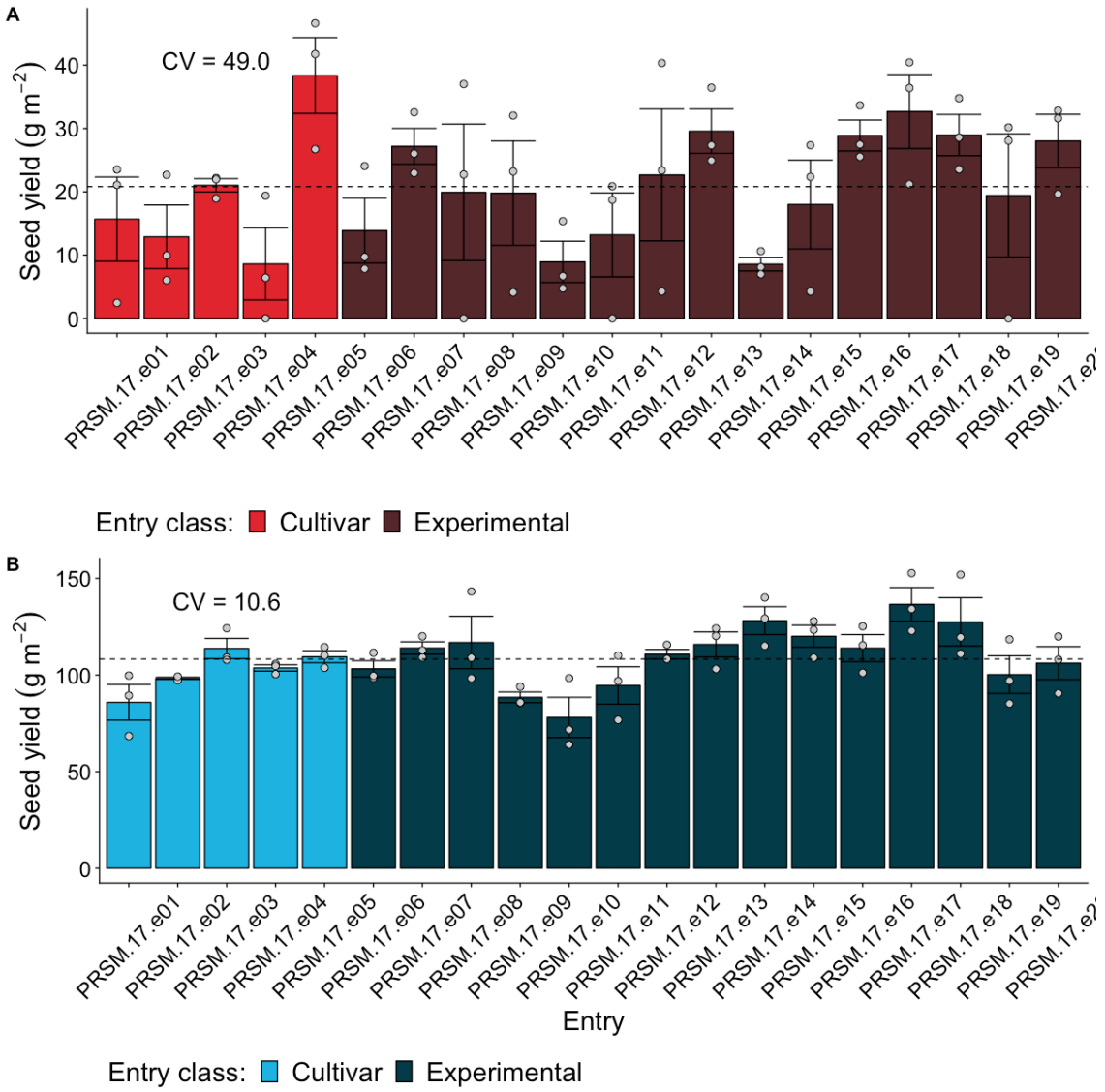
§ Winter survival as reported by Minnesota Turf Seed Council in relation to NK200 in 2016 (1 unit = 12.5 % change in tiller survival).



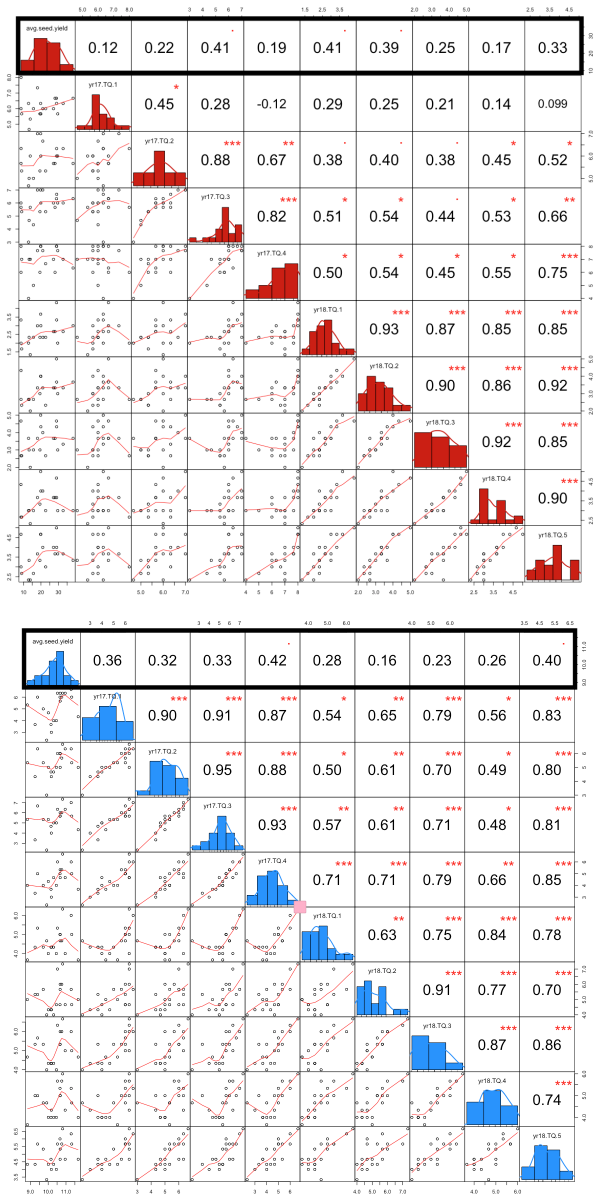
Supplemental Figure 3.1- Kentucky bluegrass seed yield plotted against turfgrass quality for 45 entries. Three groups of germplasm included were a core collection of 20 representing the USDA collection, 16 plant introduction selection with both good seed yield and quality, and 9 cultivars. Entries were grown in turf and seed environments with seed plots having three residue treatments. Data adapted from Table 3 in Johnson et al. (2003).



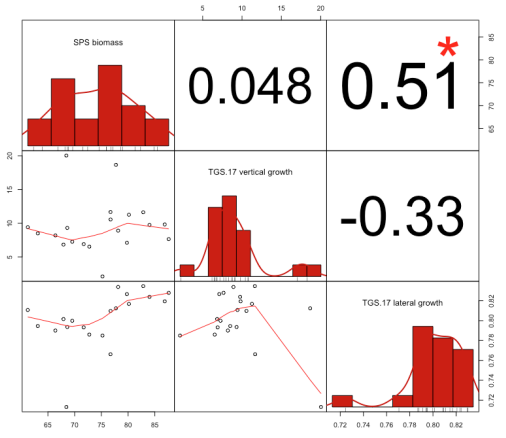
Supplemental Figure 3.2 - Diagrammatic rating scales used for measuring lateral growth and genetic turfgrass color. A) Each panel represents a 12.5% increase in green cover for each of the nine density categories. B) Custom color palette for rating genetic color. Scale values (HSB): 1 = 60-100-90, 2 = 67-100-86, 3 = 74-100-86, 4 = 81-100-62, 5 = 88=100=50, 6 = 96=100-37, 7 = 102-100-31, 8 = 111-90-26, 9 = 120-90-20.



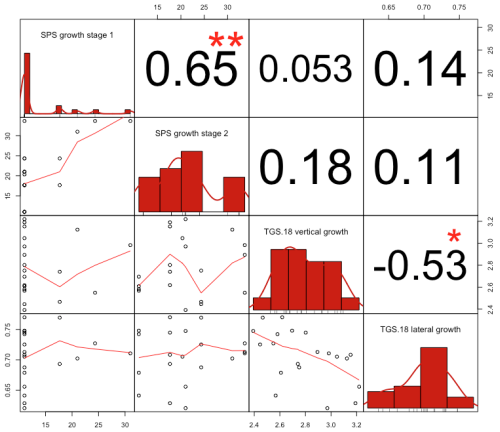
Supplemental Figure 3.3 - Seed yield (g m^{-2}) for each of the 20 entries at both locations with cultivar and experimental populations denoted with different colors. A) Roseau average seed yield was 21 (g m^{-2}) with a coefficient of variation of 58.3%, B) St. Paul average seed yield (105 g m^{-2}) with a coefficient of variation of 16.7%. Bars are equal to +/- 1 standard error.



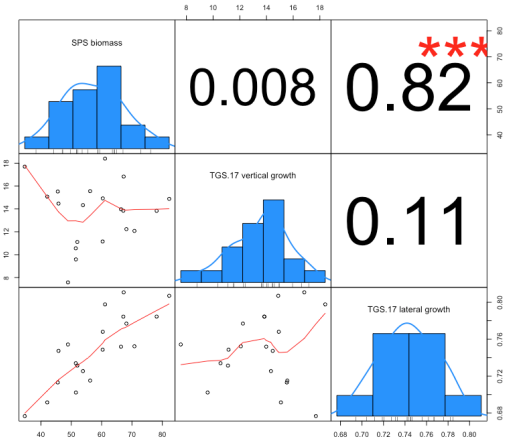
Supplemental Figure 3.4 - correlation matrices exploring all turfgrass rating dates relationship with seed yield. Black box outlines seed yield correlation coefficient with each rating date across the establishment and production year. Roseau is in red.



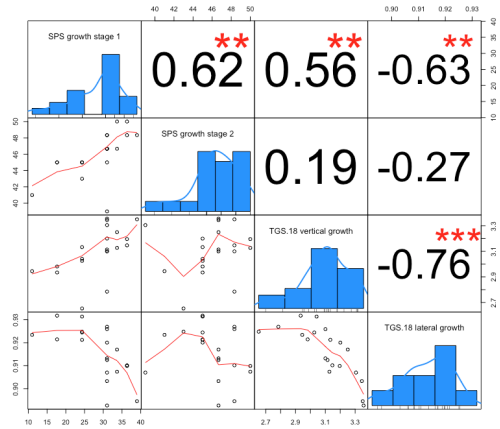
1A)



2A)



1B)



2B)

Supplemental Figure 3.5 - Pearson correlations in 2017 and 2018 for growth and maturity parameters. Panels 1A and 1B) relationships between turfgrass vertical, lateral growth rate, and seed production biomass accumulation in Roseau and St. Paul respectively. Panels 2A and 2B) relationships between turfgrass vertical, lateral growth rate, and growth stage in the seed production environment in Roseau and St. Paul respectively.

Chapter 4

Relationships and influence of perennial ryegrass yield components on total seed yield in spaced plant and sward environments

Introduction

Perennial ryegrass (*Lolium perenne* L.) is one of the most important forage and amenity grasses worldwide, providing excellent nutritional quality as well as rapid establishment and wear tolerance. Although best adapted to moderate climates, perennial ryegrass is also grown and produced for seed in colder continental regions. While nutritional and aesthetic traits define cultivar performance to the consumer, it is seed yield potential that gives rise to cultivar success (Elgersma, 1985). This has led to extensive research in seed production systems in both Europe and the United States. Major increases in seed yield have been realized through advancements in agronomics, and important consumer traits such as forage yield and turfgrass quality have been improved through breeding (Hart et al., 2012; McDonagh et al., 2016). Despite breeding efforts beginning in the early 20th century, seed yield potential of new cultivars has not been substantially improved (Casler, 2006; Sampoux et al., 2013).

Drivers of yield potential

Perennial ryegrass tillers are monocarpic and flower conditionally as function of primary floral induction in the fall and secondary floral induction in the spring. Potential seed number per unit area (or plant) is determined prior to anthesis and is the product of fertile tiller number (FTN), spikelets per spike, and florets per spikelet. Potential seed yield has typically been calculated after harvest using the measure of hundred seed weight (HSW) multiplied by potential seed number. Success of fertile tiller induction and filling of florets with ripe seed, which defines realized seed yield, is both universally and substantially less than the potential yield.

Fall growth influences seed yield in many cool-season grasses such as orchardgrass (*Dactylis glomerata* L.) and tall fescue (*Festuca arundinacea* Schreb.) (Chastain and Grabe, 1989a; Chastain and Grabe, 1989b), but not necessarily perennial ryegrass (Chastain and Young, 1998). Winter injury can cause yield reduction or even complete stand loss regardless of fall establishment, however, larger plants with well developed crowns have a higher chance of survival (Ehlke and Vellekson, 1998; Waldron et al., 1998). That being said, perennial ryegrass has the ability to compensate from some winter damage and maintain yield more than some grasses by virtue of inherent low juvenility in spring-

produced tillers (Aamlid et al., 2000), which allows spring-produced non-vernalized daughter tillers to flower. Spikelet number per spike is related to the number of primary branches on the shoot apical meristem. Apex size is determined by the timing of tiller initiation and it is likely that fall growth and winter survival play some part in the number of spikelets and subsequent florets per spike (Colvill and Marshall, 1984).

In several grass species increases in the number of fertile tillers is often indicative of increased seed number. This is not always the case in perennial ryegrass, as FTN was found to be highly influential in seed yield of spaced plants under low competition (Bugge, 1987), but not in a competitive sward environment (Chastain and Young, 1998). This is probably due to the complex compensatory nature of yield components. For example, Donald (1954) studied yield components of Wimmera ryegrass (*Lolium rigidum* Gaud.) across plant densities that varied in a logarithmic fashion (0.5 to 922 plants 400 cm^{-2}). Although FTN increased across the density gradient (41 to 590 spikes 400 cm^{-2}), fertility of individual spikes was maximized at the second lowest density (5 plants 400 cm^{-2}). At the highest density, yield per spike was reduced by over an order of magnitude (0.080 vs 0.005 g $spike^{-1}$). That being said, total yield per unit area did not change significantly between 5 and 922 plants 400 cm^{-2} (Supplemental Fig. 4.1). Donald theorized that competition reduced the individual spike fitness at very low densities through intra-plant competition and at very high densities via inter-plant competition. Seed yield components responded differently depending on the kind of competition, which may or may have led to noticeable changes in overall seed yield.

Perennial ryegrass plants produce far more opportunities for seed than are utilized leaving many empty florets. The importance of this characteristic has been described through quantification of biological floret site utilization (FSU), or the proportion of filled florets to total florets (Elgersma, 1985). Many authors have found biological FSU estimates between 50 and 70%, and of these filled florets, only a fraction make it through harvesting and cleaning (Burbidge et al., 1978; Elgersma and Śnieżko, 1988; Abel et al., 2017). Armstead et al. (2008) summarized that poor floret utilization was due to several reasons including maintenance of perenniality and resource competition leading to early seed abortion. Trethewey and Rolston (2009) found that in perennial ryegrass the spike itself was an important source for grain fill and that covering leaf and stem material post anthesis

had very little impact on seed yield. This was because perennial ryegrass plants often accumulate far more resources than are utilized by seeds (Warringa and Marinissen, 1997). Diversion of resources to secondary and tertiary tillering may be one of the reasons for this over ambitious and seemingly wasteful energy acquisition. However, it has been shown that far more energy is reserved for tillering than is ever used for seed production with or without secondary tillering (Warringa and Keuzer, 1996). Taking away from realized yield as well is the duration and uneven nature of seed growth rate across spikelets that leads to light seed and seed shattering, both resulting in losses before or during harvest (Warringa et al, 1998).

Exogenous factors such as winter stress and the interconnected nature of seed yield components have been well studied to better understand seed yield potential. Bivariate analyses have often been conducted to describe these relationships. For instance, fall establishment may be correlated spikelet number, which then may be individually related to seed yield. However, the co-dependent nature of yield components and exogenous influences from climate can make results difficult to interpret. Dewey and Lu (1959) used bidirectional correlation path analysis to measure the direct and indirect influence of seed size, spikelets per spike, spikelet fertility, and plant size on seed yield. Fertility and plant size were shown to have the largest direct effect on seed yield; however, it was remarked that seed size had a very sizable indirect effect on yield via plant size. This analysis approach, although accurately highlighting the complex nature of yield components, did not take into account the very reasonable assumption that harvested seed size does not dictate plant size. To make interpretation of results more realistic, Dofing and Schmidt (1985), suggested a unidirectional model in which the theoretical sequence of yield component development was assumed (e.g. plant size influences seed size). Measuring the interconnectedness of yield components in their influence on total seed yield sheds light on meaningful breeding targets.

Evaluation of changes in yield components of domesticated grains is a prime example from which an ideotype of a grain producing grass plant can be hypothesized (Sedgley, 1991). Philipp et al. (2018) found that spikelet number in wheat (*Triticum aestivum* L.) was not correlated with seed yield in either landrace or elite wheat lines, and that breeding had no appreciable impact on spikelet number. Results did show substantial

increases in yield per spikelet in elite compared to landrace lines. This indicated that plant breeders have been selecting for genotypes that have increased fertility (seeds per spike) and uniform grain fill within spikelet, but not potential seed number (spikelets per spike). Have breeders made a specific effort to improve fertility, or was this an indirect product through which an ideotype was *a posteriori* defined? It is interesting to note that, in this case, yield on a per unit basis was more highly correlated with yield per spikelet in landrace ($r_p = 0.43$, $P < 0.001$) than in elite lines ($r_p = 0.20$, $P < 0.01$), indicating fertility could have been easily detected as an early primary selection target by breeders. Certainly, others have utilized the ideotype of compactness and high fertility to increase yields in undomesticated crops and cereals (Sedgley, 1991; Lodge, 1993). Through the measurement of both indirect and direct effects of yield components, a stronger idea of breeding targets should be definable.

Theoretical seed yield construct

Correlation path analysis was initially devised to handle the simultaneous influence of several variables in the study environment through the use of correlation coefficients. Although path analysis is still actively being used to analyze yield components it may be advantageous to employ a more contemporary approach (Studer et al., 2008; Abel et al., 2017; Cooper et al., 2012). Structural equation modeling (SEM) is similar to path analysis in that it combines both indirect and direct effects simultaneously, however, SEM relies upon covariance measures and allows for the inclusion of latent variables (Grace, 2006). It is important to note that SEM is not built upon hypothesis testing per se, but a body of literature and an individual's experience with the study environment. In order to apply causation beyond that of the *P*-value associated with paths in the model, theoretical knowledge and experience must be applied to the analysis *a priori*. Recently, SEM has been used to determine yield components of flax (*Linum usitatissumum* L.) and oats (Zhang et al., 2014; Lamb et al., 2011), as well as weed suppression in winter oats (*Avena sativa* L.) (Fradgley et al., 2017) and cover crop impact on dry bean (*Phaseolus vulgaris* L.) yields (Hill et al., 2017). In this study, known relationships between yield components and environmental variables were used to develop a theoretical construct for perennial ryegrass seed yield in northern climates (Fig. 4.1).

Growth rate and crown width prior to fall were included as indicators of plant resources prior to dormancy and are related to winter survival of perennial ryegrass in cold climates (Hulke et al., 2007; Waldron et al., 1998) with fall produced tillers often yielding more than spring produced tillers (Colvill and Marshall, 1984). For this reason, winter survival and growth rate may both directly and indirectly influence seed yield. Once stems elongate and the spike emerges, the spike itself along with whatever energy the culm has stored becomes a potential source for grain fill (Trethewey et al., 2008). Therefore, longer spike length may positively increase seed yield on a per spike basis. Logically, length and spikelet number could have some direct influence on seed yield per unit area (Yamada et al., 2004). Surprisingly this is often not the case; in fact results sometimes show that spikelet and floret number along with FSU have no direct impact on seed yield (Elgersma, 1990b; Hampton and Hebblethwaite, 1983).

Elgersma (1990c) surmised that seed yield on a per unit area basis, can best be predicted from the number of cleaned seeds. The cascade of yield components suggests that FSU is correlated with seeds per spike and controlled through the number of spikelets per spike. The realized or economic FSU is typically around 20-50% and is the proportion of clean seed number to potential seed sites (Burbidge et al., 1978; Elgersma, A. 1990b; Rolston et al., 2007). Economic FSU is driven by the number of ripe, plump seeds and can only be measured post cleaning (Elgersma, 1985). More uniform and efficient use of florets, or simply fewer florets in the seed head, drives higher economic FSU values (Warringa et al., 1998). For this reason, economic FSU has been related to the clean seed mass to total spike biomass ratio. This proxy measurement, called fertility index, has been found to be highly correlated with FSU (Trupp and Carlson, 1971; Raeber and Kalton, 1956; Jessen and Carlson, 1985). Unlike FSU, fertility index has an upper bound far lower than 1 simply due to the weight of the anatomical building blocks associated with the spike (rachis, rachilla, glumes, etc.). A high fertility index could mean a higher economic FSU *per se* or a more compact spike, such as is found in wheat. So although yield per spike is used in the calculation of fertility index, it is hypothesized to directly influence spike yield, whereas spike yield and FTN directly influence total yield (Fig. 4.1).

Application to a breeding program:

Utilizing structural equation modeling is an interesting way to confirm the interconnectedness of yield components to yield, beyond this however, it may also have utility in developing breeding targets for seed yield. The initial unit of selection in breeding programs for outcrossing grass species is typically the spaced plant. There are several good reasons for this including ease of observation, small space and seed requirements, and ability to select within and among families (Vogel and Pedersen, 1993). However, breeding values measured in the spaced plant environment must be indicative of what will be realized in a sward environment. This is often not the case as many traits including seed yield respond differently depending on the level of competition in the growing environment (Elgersma, 1990; Waldron et al., 2008). Lazenby and Rogers (1964) determined that making selections for biomass in perennial ryegrass in wide spacings (3 plants m^{-2}) gave the greatest phenotypic plasticity compared to closer plantings (23 plants m^{-2}), unfortunately the differences found at the wide spacings were often not associated with differences found at the sward level. In fact, the authors found there could sometimes be an interaction between the top and bottom performer at wide spacings and in a sward. Perennial ryegrass breeders would benefit from research focused on not only the importance of yield components, but also interactions across selection environments.

The objective of this study was to measure seed yield and yield components concurrently in a typical seed production sward (SPSward), as well as competitive (CSPN) and noncompetitive (SPN) spaced plant environments. Relationships between spaced plant and the sward environment were explored through the establishment of a theoretical construct that was used to formulate SEMs. Results were interpreted in the context of developing potential breeding targets for seed yield improvement.

Materials and Methods

Plant material

Perennial ryegrass experimental entries consisted of 15 interrelated populations and 5 cultivars. Interrelated populations were aggregated from several half-sib families all originating from one maternal parent. The parent from each population was selected from advanced synthetic material that was tested for turf performance at the University of Minnesota and the University of Connecticut (Supplementary Table 4.1). Selections were

sent to Oregon for additional screening and crossing (DLF International Seeds). The five cultivars were chosen based on a wide range of performance for winter survival, turfgrass quality, seed yield capacity, and consisted of four turf-type and one forage-type perennial ryegrass (Supplementary Table 4.1).

Experimental locations and management

Seed production plots and spaced plant nurseries were established in June 2017. The experiment was tested at two locations in Minnesota, one at the Minnesota Agricultural Experiment Station at the University of Minnesota in St. Paul, MN and the other at the University of Minnesota Magnusson Research Farm in Roseau, MN. The two locations were located approximately 560 km apart. Roseau County is the major seed production region in Minnesota and the University of Minnesota turfgrass and forage breeding programs are located in St. Paul. The soil at Roseau was a Borup silt loam soil with 2.7% organic matter and a pH of 7.7; St. Paul had Waukegan silt loam soil with 4.8% organic matter and a pH of 6.4.

Seed production plots and spaced plant nurseries were maintained with the same management practices. Upon establishment, plots and nurseries were fertilized using a Gandy drop spreader (Gandy Company, Owatonna, MN) at a rate of 24.4 kg N ha^{-1}, 40.1 kg P ha^{-1}, and 24.4 kg K ha^{-1}. In late September plots were fertilized using a Gandy drop spreader at a rate of 0.0 kg N ha^{-1}, 33.6 kg P ha^{-1}, and 33.6 kg K ha^{-1}. Broadleaf weeds were controlled in the establishment year with a single application of bromoxynil and MCPA [(3,5-dibromo-4-hydroxybenzonitrile and 2-ethylhexyl ester of 2-methyl-chlorophenoxyacetic acid); Bison, BASF, Research Park Triangle, NC] at a rate of 0.28 kg a.i. ha^{-1} at each environment. Fall and early spring weed seedlings were controlled with a single application of pendimethalin (N-(1-ethylpropyl)-3,4-dimethyl-2,6-dinitrobenzenamine; Prowl H_2O, BASF, Research Park Triangle, NC) at a rate of 2.12 kg a.i. ha^{-1} at each environment. Fall broadleaf weeds were controlled with an application of 2,4-D (2, 4-dichlorophenoxyacetic acid; Shredder, Winfield St Paul, MN) at a rate of 0.40 kg a.i. ha^{-1}. In the seed production year fertilizer was applied in early May at a rate of 123.1 kg N ha^{-1}. Broadleaf weeds were controlled in late May with an application of 2,4-D (2, 4-dichlorophenoxyacetic acid; Shredder, Winfield, St. Paul, MN) at a rate of 0.27 kg a.i. ha^{-}

[1] and dicamba (3,6-dichloro-o-anisic acid; Sterling Blue, Winfield, St. Paul, MN) at a rate of 0.27 kg a.i. ha^{-1} at each environment. All chemical applications were applied using a CO_2-powered bicycle sprayer equipped with a 2.75 m boom and eight 1002 TurboTeeJet® (TeeJet, Springfield, IL) nozzles operating at 186 kPa in water equivalent to 117 L ha^{-1}.

Climate data for Roseau were collected from the North Dakota Agricultural Weather Network, Fox station (48.878013, -95.847207) and the University of Minnesota St. Paul weather station (44.995031, -93.185839) for both years. Growing degree days (GDD) were calculated using a base temperature of 0°C (Table 4.1). Data were taken on soil temperature 10 cm below the soil surface.

Experimental environments

Seed production swards were seeded with a Hege 1000 small-plot seeder (Hege Equip. Inc; Colwich, KS). Each plot was 3.6 m long consisting of 10 drilled rows 15 cm apart. Each plot was seeded with 6.5 kg ha^{-1} of pure live seed resulting in approximately 70 seeds per linear meter of row. Each location had three blocks, along with seeded border plots, arranged in a randomized complete block design.

Plant material for SPN and CSPN were seeded in April and allowed to grow in the greenhouse until transplanted into the field in the first week of June. Standard non-competitive spaced plant nurseries were arranged in a 20 x 20 Latin square with three squares (blocks). Plants from each entry were randomized within each row and column and were planted 0.61 m apart (3 plants m^{-2}), which is a typical distance that minimizes competition. Competitive spaced plants were transplanted into units that facilitated moderate competition and consisted of 10 plants of the same entry. Competitive units were aggregated so that each block contained 200 plants (20 entries x 10 plants). The CSPN environment included six blocks instead of three to minimize localized winterkill (Supplementary Fig. 4.2). Plants in CSPN were planted 0.21 m apart (23 plants m^{-2}), which fostered competition, but allowed observation on spaced plants (Lazenby and Rogers, 1964). Both CSPN and SPN environments were represented by 1200 plants at each location (Supplementary Fig. 4.3).

Seed production, SPN, and CSPN environments were arranged in a randomized complete block design within location (Supplementary Fig. 4.2). This was accomplished

by randomizing growing environments at the block level. Blocks were arranged in this way so that statistical comparisons could be made between entries across growing environments.

Data collection

Stand counts were made in SPSward at both locations 14 d post seeding on two 0.5 m transects of row. Biomass was collected in the field as fresh weight in late August 2017 on both spaced plants and three-0.5 m transects in seed production plots. Crown width was only measured for spaced plants and was determined by measuring the width at the base of each plant following biomass collection. Winter survival was measured through repeated visual estimates of spring green-up on a 1-9 scale (Hulke et al., 2007). On a spaced plant basis, a 5 would describe 50% of the tillers on a plant healthy and green. Green-up in seed production plots was estimated from the proportion of tillers that were showing active growth relative to the size of the plot minus any previous data or sign on poor stand quality prior to winter. Because living tillers may or may not have been noticeably growing or green during the initial rating and new tiller emergence begins to confound later ratings, an average of two ratings two weeks apart was considered winter survival.

Sampling for seed yield and yield components occurred in early July in St. Paul and early August in Roseau. Two harvests were made in St. Paul and three were made in Roseau to ensure seed was ripe at the time of harvest. Spaced plants and seed production swards each had three samples collected from each experimental unit.

Seed production sward yield components were determined by: 1) Sampling four-10 spike samples from each plot to determine individual spike weight, spike length, spikelet count, seed weight and spike seed yield. Spike samples were kept safe from shattering and breakage by storage in 10 x 4 x 40 cm plain unwaxed paper bread bags (Bagcraft Papercon). 2) Fertile tiller number samples were taken by cutting four-15 cm transects along a row within each plot so that all culms were exposed, and a flush surface was created, followed by a second cut that was made so that a 3-4 mm segment from each culm dropped into a bin. These segments were funneled into an envelope to be counted later. 3) Whole plot yield (seed yield) was measured collecting two 1 m^2 swaths of perennial ryegrass per experiential unit at approximately 400 g kg^{-1} moisture content.

Spaced plant seed yield and components were determined for each plant in each nursery by: 1) Sampling one 10-spike samples from each plant. These spike samples were also stored in plain unwaxed paper bread bags to maintain spike integrity. 2) Fertile tiller number samples were taken in the same fashion as plot transect except all the culms on a plant were made flush and then, depending on the size of the plant, two cuts were made so culm pieces could be collected. 3) All plant material left after fertile tillers and spike samples were collected was collected into a paper bag to measure whole plant yield. Sample identities were maintained using a barcoding system.

All samples were dried for 5 d at 35 °C in a forced air dryer. Sample 1 was used to determine 10 spike mass and seed yield, of these, four representative spikes were chosen to measure spike length and count spikelet number. Seed from the 10 spikes was counted using a Seed Count S-JR seed counter (Data Technologies, Kibbutz Tzora, Israel) to determine thousand seed weight. Floret site utilization was not measured directly but approximated through the ratio of spike seed weight to total spike biomass. Sample 2 was used to calculate FTN where culm pieces were separated from leaf and inert material using setting 15 on a General Seed Blower (Seedburo Equipment Company, Des Plaines, IL). Fertile tiller number was assessed by counting stem pieces using a Seed Count S-JR seed counter (Data Technologies, Kibbutz Tzora, Israel). A strong positive relationship was found between samples that were hand and then machine counted (r_s = 0.96, $P < 0.001$). Sample 3 was used to determine seed weight on a per plot and plant basis; each sample was processed in a LD350 thresher with a 2.5 x 7.5 mm concave and ran at 800 rpm (Wintersteiger Corp., Ried im Innkreis, Austria). Following threshing, seed and chaff were separated by passing the processed seed through a 1.625 x 9.525 mm sieve. light seed was removed by aspiration using a Superior® Fractionating Aspirator (Carter-Day Int. Inc., Minneapolis, MN).

SEM construct

The theoretical construct is displayed in Figure 4.1 and was used as the baseline model for each growing environment at each location. Models for the SPSward environment had additional paths added beyond the confines of the initial construct for model fitting purposes (see statistical analysis). Each path has an alpha-numeric code

assigned to aid in describing the theoretical direct and indirect implications of the path on seed yield.

A single latent variable was included in the model for plant resources, which was approximated through fall indicator measurements of crown width and biomass accumulation. Latent variables represent a concept and are approximated by one or several indicator variables The amount of plant resource accumulation the previous fall could offer some potential indication as to the potential for the plant to survive the winter (Hulke et al., 2007; Waldron et al., 1998) and biomass production the year preceding the production year has been indicative of seed yield potential in other grasses (Chastain and Grabe, 1989a; Chastain and Grabe, 1989b).

Winter survival was coded to directly influence spikelet per spike (Fig. 4.1-1A), FTN (Fig. 4.1-1B), and total seed yield (Fig. 4.1-1C). In this way, winter survival was allowed to influence seed yield through all possible pathways. Winter conditions in Minnesota can completely kill a ryegrass stand leading to no or greatly reduced yield (Fig. 1-1B and C) (Elling, 1979; Elling, 1980; Heineck et al., 2018). Furthermore, fall produced tillers are often larger and produce more seeds than spring produced tillers (Fig. 4.1-1A) (Colvill and Marshall, 1984). Winter survival was hypothesized to have a direct positive relationship with all variables.

Perennial ryegrass spikes are determinate and potential spikelets per spike are formed prior to culm elongation with summer and autumn produced tillers developing more spikelets and longer spikes than spring produced tillers (Fig 4.1-2A) (Hill and Watkin, 1975). The number of spikelets per spike was hypothesized to be positively linked to seed yield through yield per spike but has generally not been directly correlated with yield (Elgersma, A. 1990b). Spikelet number was hypothesized to be directly and positively influence spike length and fertility index (Fig. 4.1-2A and B). Fertile tiller number can be seen as both theoretically contributing to and detracting from seed yield. Fertile tillers have been shown to be directly positively correlated with seed yield in both swards and spaced plants (Bugge, 1987) (Fig 4.1-3C). However, too many fertile tillers may lead to increased competition thereby lowering resources on a per spike basis (Donald, 1954; Supplementary Fig. 4.1) (Fig. 4.1-3A and B).

Spikes are a photosynthetic source for seeds and increased length may increase seed yield per spike (Fig. 4.1-4A), however, an increase in length would decrease fertility index by increasing spike weight (Fig. 4.1-4B) (Trethewey et al., 2008). Seed weight generally impacts FSU and therefore was positioned to positively influence fertility index (Fig. 4.1-5) (Colvill and Marshall, 1984). Fertility index measures the relative yield per spike in relation to total spike mass (Jessen and Carlson, 1985). Because of this there will likely be a positive relationship between fertility index and spike yield (Fig. 4.1-6). Yield per spike along with FTN is theoretically a major direct contributor to total yield (Fig. 4.1-7).

This network of paths would, in theory, limit the effects of some yield components through indirect relationships. For example, spikelet number had a theoretical positive indirect effect on yield through fertility index (Fig. 4.1 2B➤ 6➤ 7), but a negative indirect influence through spike length may also reduce its total effect (Fig. 4.1 2A➤ 4B➤ 6➤ 7).

Statistics

Analysis of all data was conducted in the R environment (R Core Team, 2018). Analysis of variance for sward establishment, winter survival and seed yield on both a spike and experimental unit basis was conducted using a linear mixed effects model in the 'lme4' package (Bates et al., 2015). Location and entry were modelled as fixed effects. Blocks, along with rows and columns in the SPN environment were modelled as random effects. Marginal means were estimated on an entry basis and Tukey's HSD was employed to conduct means comparisons ($\alpha = 0.05$) using the 'emmeans' package (Lenth, 2018). Correlations for a single trait between growing environments were conducted on an entry mean basis using Spearman's rank correlation estimation. All graphical analysis was conducted using ggplot (Wickham, 2016). Tests for significance between growing environments for individual traits were conducted with a two-sample paired t-test on an entry mean basis.

Structural equation modeling was conducted using the 'lavaan' package (Yves Rosseel, 2012). Prior to model fitting a theoretical construct was developed (Fig. 4.1) (Grace, 2006). All data were checked for normality and scaled prior to analysis to induce equal variance among variables using the default setting in the scale() function. Paths were only added to models that were founded both on personal experience, causal assumption,

and the literature. Unadjusted means were used to fit the SEM model. In this way, the modeling approach maintained a confirmatory rather than an exploratory nature. The SEM construct was designed with eight observed variables and one latent variable for plant resources. Initial model fit was assessed using the comparative fit index (CFI), root mean square error of approximation (RMSEA), and chi-square tests. The chi-square goodness of fit is used to compare model implied vs observed variance-covariance components, and so insignificant values are preferable. Due to the extremely large sample size of the spaced plant environment (n = 1200) a significant chi-square statistic was deemed acceptable to avoid overfitting. However, a non-significant value was reached ($\alpha = 0.05$) for the SPSward environment by adding additional paths. Paths were considered based on modification index values and their theoretical biological implications. For instance, a path that implied seed weight influenced FTN was deemed unreasonable. Only significant path coefficients were considered non-zero in calculation of indirect and direct effects. Due to the large discrepancy number of experimental units the alpha was set at 0.10 and 0.001 for SPSward (n = 60) and spaced plant environments (n = 1200) respectively.

Results

Establishment and winter survival

Seed production sward stand counts at Roseau averaged 32 plants m row^{-1} and St. Paul avenged 49 plants m row^{-1}, however no differences were found between entries within location. Spaced plant nurseries suffered very minor losses in plant numbers in the establishing year due to insect or drought. Growing degree day units and precipitation in the establishing year were slightly greater in St. Paul compared to Roseau (Table 4.1). This resulted in generally more biomass being produced in St. Paul in the initial growing season, except in SPSward (Table 4.2).

Winter conditions in Roseau brought about lower minimum soil temperatures and far more days of sub zero soil and air temperatures compared to St. Paul (Table 4.1). This along with low amounts of precipitation in spring 2018 probably led to the higher winterkill at Roseau across growing environments (Table 4.2). The SPSward and CSPN environments had similar levels of winter survival, whereas the non-competitive plants had far lower survival (Fig. 4.2). In Roseau about 60% of CSPN and SPSward tillers survived

compared to only 30% in the SPN environment. Analysis of variance for winter survival found differences between entries at both locations and in all growing environments ($P < 0.05$). In the seed production year, growing degree day accumulations and precipitation were both much higher in St. Paul compared to Roseau.

Trends in seed production

Seed yield was measured both on an experimental unit and per spike basis. Analysis of variance for total seed yield showed that there were differences between entries at both locations and all three growing environments ($P < 0.05$). Seed yield in St. Paul were double of those in Roseau across all growing environments, however spike yield varied little (Table 4.2). Sward production seed yields did not consistently correlate strongly with any seed yield component variable (Fig. 4.5). At Roseau SPSward plots that had high FTN also had higher yields ($r_p = 0.77$, $P = < 0.001$), but this relationship was not as strong in St. paul ($r_p = 0.35$, $P = 0.01$). Fertility index too was related to seed yield within the SPSward environment at both Roseau and St. Paul ($r_p = 0.67$ and 0.28, $P < 0.001$ and 0.03). Winter survival was correlated to seed yield in Roseau ($r_p = 0.88$, $P < 0.01$), but not St. Paul. Seed yields in St.Paul were lower per plant in CSPN than in SPN ($P < 0.001$) and there was very little difference between spaced plant yields in Roseau, although competitive plants were slightly higher yielding ($P = 0.09$). Mean spike yields between SPN and CSPN were similar at both environments.

SEM: model fit

Six SEMs were built using the hypothesized construct (Fig. 4.1). All model CFI values were above 0.91 indicating a reasonable fit (Hu and Bentler, 1999). RMSEA values were also acceptable, ranging between 0.09 and 0.16. However, chi-square test statistics were all significant ($\chi_{20}^2 < 0.001$). Additional paths were added to SPSward models beyond those included in the original model to improve model fit. Four additional paths were added to both the St. Paul and Roseau SPSward models (Supplemental Fig. 4.4). It is important to note that the addition of these paths drastically changed the chi-square statistic for both St. Paul and Roseau ($P = 0.05$ and 0.06 respectively) but were nearly inconsequential to the results.

SEM: SPSward results

SEM results for seed yield can be most easily interpreted by viewing Fig. 4.4, however all individual path coefficients are included in Supplemental Fig. 4.4. Standardized indirect and direct path coefficients are presented in terms of the influence of change on total yield (plant or plot) in units of standard deviation (σ).

In Roseau, SPSward seed yields were driven almost exclusively and directly by winter survival (Fig. 4.4). Winter survival had significant indirect effect on seed yield through FTN ($\sigma_{indirect} = 0.19$, $P < 0.001$), but the vast majority of influence was direct ($\sigma_{direct} = 0.69$, $P < 0.001$). Fertile tiller number had a highly significant direct effect on seed yield ($\sigma_{direct} = 0.30$, $P < 0.001$), and no indirect effects were found. Spike yield had no importance in the Roseau SPSward SEM despite having a significant bivariate correlation with plot yield ($r_p = 0.36$, $P = 0.01$). Because of this, fertility index for spaced plant environments had no indirect impact on seed yield but was highly influential on per spike yield ($\sigma_{direct} = 1.30$, $P < 0.001$).

In St. Paul, SPSward seed yields were not correlated with winter survival (Fig. 4.5) despite significant variation existing between entries for this trait ($P = 0.02$). Fertile tiller number had the greatest influence on seed yield of any trait ($\sigma_{direct} = 0.31$, $P < 0.001$), with no significant indirect effect. However, both fertility index and seed yield per spike had a significant influence almost as large as FTN (Fig. 4.4). Fertility index and spike length were significant contributors to spike yield ($\sigma_{direct} = 0.95$ and 0.70 respectively, $P < 0.001$) (Supplemental Fig. 4.4). Spikelet count also had a small positive indirect effect on total seed yield ($\sigma_{indirect} = 0.10$, $P = 0.04$). This indirect relationship was very complex, and was regulated by a direct positive relationship through fertility index ($\sigma_{direct} = 0.41$, $P = 0.02$) and spike length ($\sigma_{direct} = 0.73$, $P < 0.001$) with spike length then positively influencing seed yield per spike ($\sigma_{direct} = 0.70$, $P < 0.001$) (Supplemental Fig. 4.4). Driving the generally low indirect relationship with seed yield was the negative relationship between spike length and fertility index ($\sigma_{direct} = -0.69$, $P < 0.001$), which severely reduced the total indirect influence of spikelet count on seed yield.

SEM: Spaced plant results

Competitive and non-competitive spaced plant environments had strikingly similar path coefficients within locations (Fig. 4.4). Across locations yield on a plant basis was influenced most notably by winter survival and a direct effect of FTN. At Roseau, the influence of winter survival had only an indirect influence on SPN plant yield ($\sigma_{indirect}$ = 0.59, $P < 0.001$), but both an indirect ($\sigma_{indirect}$ = 0.48, $P < 0.001$) and direct (σ_{direct} = 0.33, $P < 0.001$) effect on CSPN plant yield. The effect of FTN was much greater than any other variable in SPN nurseries and there was little influence of fertility index or spike yield by comparison (Fig. 4.4). Competitive spaced plant yield, however, was largely influenced by both fertility index and spike yield. It is interesting to note that fertility index and plant yield in SPN and CSPN nurseries had large bivariate correlation coefficients (r_s = 0.53 and 0.67, $P < 0.001$ respectively) .

At St. Paul, there were almost no differences between indirect and direct path coefficients for CSPN or SPN. Winter survival and FTN were the most important in both growing environments (Fig. 4.4); however, CSPN were more influenced by fertility index and spike yield in relation to the total effect of winter survival and FTN. Competitive spaced plants also had a higher direct effect of winter survival than the SPN environment (σ_{direct} = 0.31 vs 0.22 respectively, $P < 0.001$). Yield components were more highly affected by winter survival in SPN, which increased the indirect effect disproportionately compared to the CSPN environment (σ_{direct} = 0.46 vs 0.32 respectively, $P < 0.001$). Similar to SPSward, spike length had a consistent positive influence on spike yield, but a negative influence on fertility index.

Predictive ability of spaced plants

Plot yield in the SPSward environment was compared to spaced plant and spike yields on an entry basis. No relationship was found between SPN whole plant yields and SPSward yields at either Roseau or St. Paul ($P > 0.05$). Competitive nurseries were positively correlated with SPSward yields at Roseau (r_s = 0.66, $P < 0.001$) and St. Paul (r_s = 0.64, $P < 0.001$) (Fig. 4.2). There was no relationship between spaced plant FTN and SPSward plot yields at either location ($P > 0.05$). Mean yield per spike varied very little across locations or growing environments (Table 4.2). Significant correlations between spaced plant and SPSward per spike yield were found for both CSPN and SPN nurseries in

St. Paul (r_s = 0.56 and 0.57 respectively, $P < 0.001$), however no relationship was found in Roseau. Fertility index was highly variable across entries and ranged from 0.27 in Roseau SPN to 0.49 in St. Paul SPSward environment. Fertility index values in both spaced plant environments were highly correlated with SPSward yield at St. Paul, but only the CSPN design was significant in Roseau (Fig. 4.3).

Discussion

Perennial ryegrass seed producers require high yielding cultivars to both maximize profits and meet consumer demands. One barrier to increasing seed yield in perennial ryegrass is the relative importance of this trait compared to and potential tradeoffs with forage or turf quality (Casler, 1996). Furthermore, seed yield is highly complex and yield components are both onerous to measure and difficult to interpret. Rapidly increasing seed yield in perennial ryegrass through breeding requires identification of critical yield components in order to target meaningful traits. Furthermore, although the straightforward selection of yield *per se* will likely be effective over time, the development of a yield component ideotype may be more efficient in the long term. For example, Lodge (1993) selected for seed yield in wallaby grass (*Danthonia* spp.) and found increases in per spike yield were associated with increases in inflorescence compactness and width of glume. In this case direct selection for seed yield may have encouraged a positive indirect phenotypic change. However, selecting for yield alone may inadvertently affect other traits (e.g. time to flowering) which have been associated with negative consequences for both turfgrass and forage quality (Hazard et al., 2006). Selection for higher yield in wheat has been shown to decrease competitiveness leading to increased vulnerability to weed invasion (Vandeleur and Gill, 2003).

Highly variable climate conditions at each location made combining data into a single SEM for each growing environment inappropriate. Soil temperatures at Roseau were below 0°C for over five months with extreme minimum temperatures lower than at St. Paul. These conditions led to extremely low winter survival at Roseau (Fig. 4.2) compared to St. Paul, although plants in St. Paul were still heavily damaged from winter with at least 30% of tillers succumbing to winter kill across growing environments. These results emphasize the importance of breeding for winter hardiness in perennial ryegrass in

northern climates and also facilitates an opportunity to understand its influence on yield components.

The interconnected nature of yield components can limit interpretation, but SEM allowed for simultaneous measurement of yield components and environmental conditions (Fig. 4.1; Lamb et al., 2011). Seed yield was moderated by many variables across locations and growing environments. The most prominent variable was usually winter survival with the exception of SPSward environment in St. Paul, which also happened to have the highest tiller survival. In Roseau, where tiller survival was low, results showed that winter survival had either the highest direct influence or was significantly modifying other traits such as FTN (Fig. 4.4). The direct effects of substantial winterkill on seed yield are clear; however low juvenility allows for compensation through FTN. Indirect effects of low winter survival could manifest through reduced photosynthetic potential in spring when energy is needed to promote tillering and flower induction. Although perennial ryegrass plants can compensate for some winter injury, as was seen in St. Paul, severe winter kill both lowered seed yield and shifted which variables influenced total seed yield (Table 4.2). Total yield in SPN was indirectly influenced by winter survival compared to the CSPN environments across locations (Fig. 4.4). The large indirect effect was driven mostly through a positive influence on FTN. At Roseau, spaced plants under moderate competition were impacted directly by winter kill similarly to those in the sward.

Fertile tiller number consistently had a significant indirect and/or direct effect across environments and locations. Direct effects on total seed yield were positive and significant. Indirect effects through HSW and spike yield were negligible. In fact, the direct effect of FTN on spike yield was negative in the spaced plant environment (Supplemental Fig. 4.4) but was masked by the general positive influence on seed weight. This suggests that the hypothesized negative effect may be present to a certain extent (Fig. 4.1). The positive direct effect on HSW was most likely due to larger, healthier plants having more even maturity. In the context of intra-plant resource partitioning it could be possible that FTN never reached high enough numbers to be a limiting agent on individual spike fitness.

Spikelets were hypothesized to positively influence spike length and floret site utilization directly and per spike yield indirectly. Abel et al. (2017) used correlation path analysis to find that the total effect of spikelet number and spike length were much greater

than percent seed set in perennial ryegrass. This is surprising and somewhat contradictory to was found it this study. It was interesting to observe that spikelet count had a slight positive indirect influence on seed yield in SPSward in St. Paul (Fig. 4.4) because bivariate analysis found a negatively correlated with plot yield (r_s = -0.13, $P > 0.05$). Spikelet count also had an indirect positive relationship with plant yield in both spaced plant environments. Spike length had a negative influence on fertility index, but a positive influence on spike yield. This particular finding was also masked from the bivariate analysis (Supplemental Fig. 4.5). These results provide evidence that spikelet count, if increased, could possibly increase yield but come at the cost of reduced fertility index. This would mean higher investments in spike mass for only little larger seed output. Longer spikes seemed to improve individual spike yields but had little indirect influence on seed yield due to the consistent negative relationship with fertility index. This conclusion would also have been missed using bivariate analysis and supports the hypothesis that the spike is a source for seed fill in perennial ryegrass (Trethewey et al., 2008).

Floret site utilization is a common measure of the grain site filling potential and is the link between potential and realized yield. For instance, theoretical perennial ryegrass grain yields have been shown to be nearly 10,000 kg ha^{-1}, however because economical FSU is between 20-30%, yield rarely exceeds 2500 kg ha^{-1} (Abel et al., 2017). There are many reasons for this discrepancy including maintenance of perenniality and uneven grain fill across and within spikes as well as the nature of the equation used for theoretical yield. In any case, to maximize resource use, all florets within a spike would ideally produce a seed. Elgersma (1985) explains that increasing FTN is not an acceptable breeding target fir seed yield as quality is reduced, instead efficiency of seed fill should be maximized. In this hypothetical situation, it must be made clear that the spike itself is a source for seed fill and therefore it is likely that FTN per unit area must be reduced to increase resources for individual spikes and their daughter tillers. This of course is the embodiment of harvest index in which a higher value is related to more favorable performance (Donald, 1968). In comparison to annual wheat, perennial ryegrass must maintain vegetative tillers which will always limit harvest index. However, fertility index measures the efficiency of the spike itself and is highly correlated with FSU. Wheat spikes are incredibly efficient at filling florets with harvestable grain and the measurement of fertility index has been shown to be

highly correlated with wheat yield but have a much higher heritability (Martino et al., 2015; Mirabella et al., 2016). Alonso et al. (2018) found that selection for spike fertility index was an effective means of increasing grain yields in wheat. However, the authors hypothesized that selecting grain number instead of grain mass per spike may lead to a negative influence on grain weight indirectly.

Selections based on yield data from a sward environment have been shown to provide a fairly precise estimate of production scale yield and also accommodates selection among populations or families. However, the sward environment is greatly limited by both the amount of seed required per experimental unit and the inability to measure the individual performance of genotypes within families. To make selections both within and between families breeders typically employ spaced plant growing environments and assume the mean performance on an entry basis is an accurate proxy for sward behavior. Eliminating competition between plants allows for straightforward observation of an individual but comes at a cost of vastly altering the growing conditions of the germplasm being tested. For example, plants under varying levels of competition have been known to grow and partition yield components differently (Donald, 1954), which leads to interactions between germplasm performance across growing environments (Elgersma, A. 1990d; Elgersma et al., 1994). This is quite problematic when spaced plants are the initial unit of selection in most cool-season grass breeding programs (Vogel and Pedersen, 1993). In an effort to address this discrepancy, this study was designed to explore the relationships between seed yield and its components in both spaced plant and sward environments. Furthermore, both standard non-competitive and competitive spaced plantings were tested to determine whether moderate inter-plant competition changed yield component partitioning.

In this study, SPSward were considered the gold standard to entry rankings. However, it would be useful if a less time-consuming trait could be used as a proxy for predicting performance between the two environments. Results showed that no trait within SPSward environment was a consistently strong indicator of plot yield including FTN or yield per spike yield (Fig. 4.5). In Roseau, winter survival could have been used (r_s = 0.84, $P < 0.84$), but there were very few differences between entries for winter survival at Roseau ($F_{19,38}$, P = 0.03) or St. Paul ($F_{19,38}$, P = 0.03) with non-significant rank correlations on an

entry mean basis between locations. There were few differences between entries for FTN in the SPSward environment across locations ($F_{19,76}$, P = 0.04). This would indicate that although this seed yield component was important in the sward environment it is not a sound breeding target even when selecting at the plot level. Although there was a highly significant effect of fertility index in the SPSward environment ($F_{19,74}$ = 3.2, P < 0.001), with no interaction between entry and location it did not strongly relate to plot yield at either Roseau (r_s = 0.69, P < 0.001) or St. Paul (r_s = 0.39, P = 0.02). It could certainly be the case that targeting fertility index could substantially improve seed yield, however, comparisons with spaced plantings should certainly be conducted based on plot yields.

Increasing competition between spaced plants has received mixed reviews from breeders working in perennial ryegrass. Sedcole and Clements (1973) argued against the use competitive nursery designs due to relatively high correlations between competitive (100 plants m^2) and non-competitive environments (3 plants m^2) and the difficulty of measuring traits in their competitive design. Others have also noted that the difficulty of selection within family increases when population density increases over 19 plants m^2, however, correlations between sward and spaced plant entry means have been shown to decrease without any competition (Lazenby and Rogers, 1964; Hayward and Vivero, 1984). In this study the CSPN plant population was 23 plants m^{-2} so although plants were only under moderate competition, data could easily be collected on individuals. This increased competition led to much stronger rank correlations between SPSward and CSPN compared to SPN environments. Seed production plot yields were strongly correlated with CSPN plant yields at both Roseau (r_s = 0.74, P < 0.001) and St. Paul (r_s = 0.69, P < 0.001), but not in SPN environment. These correlations were very encouraging, but it would be useful to not harvest whole plants, but instead rely upon a proxy measure. From a practical standpoint whole plants are difficult to harvest, transport, and process.

Correlations between FTN in both spaced plant environments and the SPSward environments were insignificant. This is an interesting finding as FTN was always highly important either directly or indirectly in yield all of the SEMs. FTN was highly influential on an experimental unit basis; however, there were almost no differences between entries at the SPSward growing environment. Spike yield was important in all SEM except for Roseau SPSward. Rank correlations between SPSward and spaced plants were somewhat

agreeable, with significant rank correlations existing at St. Paul for both CSPN and SPN at St. Paul (r_s = 0.67 and 0.59 respectively). There are several clues as to why spike yield is does not correlate well with plot yield. First, the direct effect of spike yield on total yield within growing environment (plot or per plant) does not have a large direct influence on seed yield, especially compared to FTN. Second, FTN and spike length often had a negative indirect effect on spike yield. These association of these two metrics with spike yield are probably implying that although a spike may yield a great deal it does not mean that is was efficiently producing grain in relation to the number of fertile tillers or size of spikes. Fertility index takes into account the mass of the spike itself and so larger numbers are more favorable. All growing environments had larger fertility indices in St. Paul than in Roseau with SPSward averaging a fertility index of 49% and both spaced plantings averaging 40%. Significant correlations between SPSward and CPSN for fertility index were found at both locations (Fig. 4.3). Fertility index includes spike yield in its calculation and so these are highly correlated. Within the context of the SEM, the indirect influence of fertility index was at least as prominent as spike yield, but probably more accurately reflects the negative influence of many large spike with relatively few seeds. This is a good sign that, similar to wheat, fertility index should be a good selection target and can also be a strong proxy for yield in sward environments.

Conclusion

Producing adequate seed yield is essential for cultivar success in both the perennial ryegrass turf and forage industries. This study examined the importance of seed yield components in both spaced plantings and seed production swards at two locations in Minnesota. Competitive (23 plant m^2) and non-competitive (3 plant m^2) spaced plant nurseries were tested. Results showed that when tiller survival was low winterkill was highly influential on seed yield both directly and indirectly through fertile tiller number. Fertile tiller number was more important in spaced plant environments than in swords, but very few differences were found between entries. Spike fertility directly influenced spike yield and was indirectly important for total seed yield. Although the relative importance of seed yield components was similar for both nursery types the competitive design had a superior predictive ability for sward yield via total plant yield and spike fertility.

Tables

Table 4.1 - Growing conditions over the course of the experiment at St. Paul and Roseau, Minnesota. At both locations 2017 was the establishment year and 2018 was the seed production year. Time periods for each season: Summer 2017: June 1st - September 30th; Winter 2017-18: October 1st - April 15th; Summer 2018: April 16th - August 15th.

	Summer 2017		Winter 2017-18			Summer 2018	
	GDD	Precip (cm)	Min Soil °C	Soil d < 0 °C	Air d < 0 °C	GDD	Precip (cm)
Roseau	2043	33.8	-4.7	155	164	1973	18.2
St. Paul	2406	40.0	-2.2	110	134	2487	36.8

GDD = growing degree days with a base temperature of 0°C

Table 4.2 - Summary statistics of traits used for structural equation modeling. The mean and standard deviation is presented for seed production swards (SPSwards), spaced plant nurseries (SPN), and competitive spaced plant nurseries (CSPN) at both Roseau and St. Paul.

Location/ Environment		Crown Width (cm)		Biomass (g)		Winter Survival (1-9)		Spike Yield (g spike^{-1})		Plot/Plant Yield (g exp. unit^{-1})	
		mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
Roseau	SPSward	-	-	74.5	19.2	2.4	1.2	0.03	0.01	20.8	11.7
	SPN	7.9	1.8	43.5	26.3	1.5	1.1	0.04	0.03	2.7	4
	CSPN	7.7	1.5	30.9	19.2	2.4	1.8	0.04	0.03	3.3	3.8
St.Paul	SPSward	-	-	58.2	18	6.3	1.3	0.04	0.01	108.3	18
	SPN	8.7	1.6	85.4	48.4	3.8	1.9	0.05	0.03	10.7	8.6
	CSPN	7.7	1.6	50.1	34.1	4.9	2.6	0.05	0.03	6.6	5.4
		Fertility Index		Fertile Tiller Ct. (ct.)		HSW (g)		Spike Length (cm)		Spikelet Ct. (ct.)	
		mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
Roseau	SPSward	0.34	0.09	32	16	0.11	0.02	12.3	1.7	14	1.6
	SPN	0.27	0.13	76	88	0.13	0.03	10.8	3.4	13.7	2.9
	CSPN	0.34	0.15	121	101	0.13	0.03	11.9	2.9	14	2.7
St.Paul	SPSward	0.49	0.07	144	25	0.13	0.01	13	2.1	14.4	1.5
	SPN	0.4	0.15	222	131	0.14	0.03	12.1	3	14.8	3
	CSPN	0.4	0.15	151	97	0.14	0.03	12.4	2.9	14.7	2.8

Figures

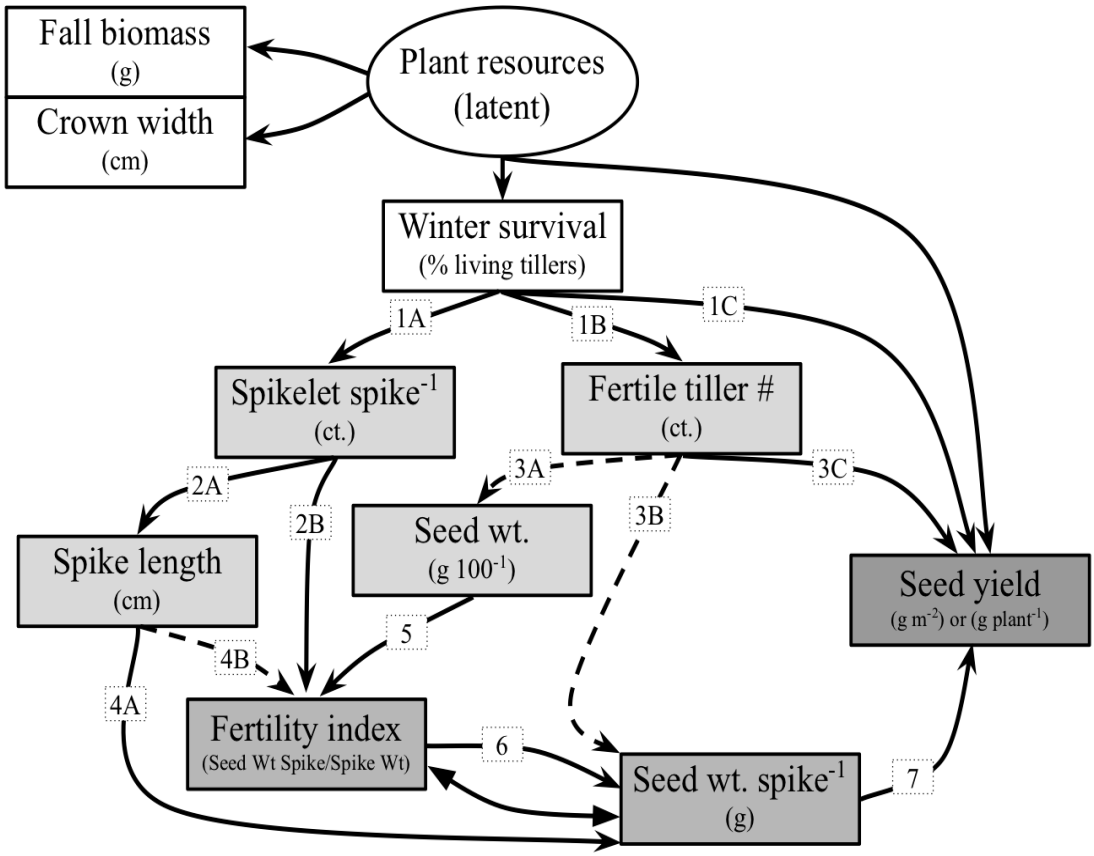


Figure 4.1 - Theoretical seed yield construct. Shaded regions represent distinct groups of variables that can influence seed yield (darkest grey box) either directly or indirectly. A) White boxes define the indirect and direct effects of the latent variable plant vigor and winter survival on seed yield. Rectangular text boxes represent observed variables, whereas the circular text box indicates the latent variable. B) Light grey boxes define the indirect and direct effects of spikelet count, spike length, seed weight and fertile tiller number on seed yield. C) Dark grey boxes define the indirect and direct influence of fertility index and seed yield per spike. Solid black arrows indicate hypothesized positive direct paths, whereas dashed arrows indicate negative paths. Double headed arrow indicates shared error variance between two variables.

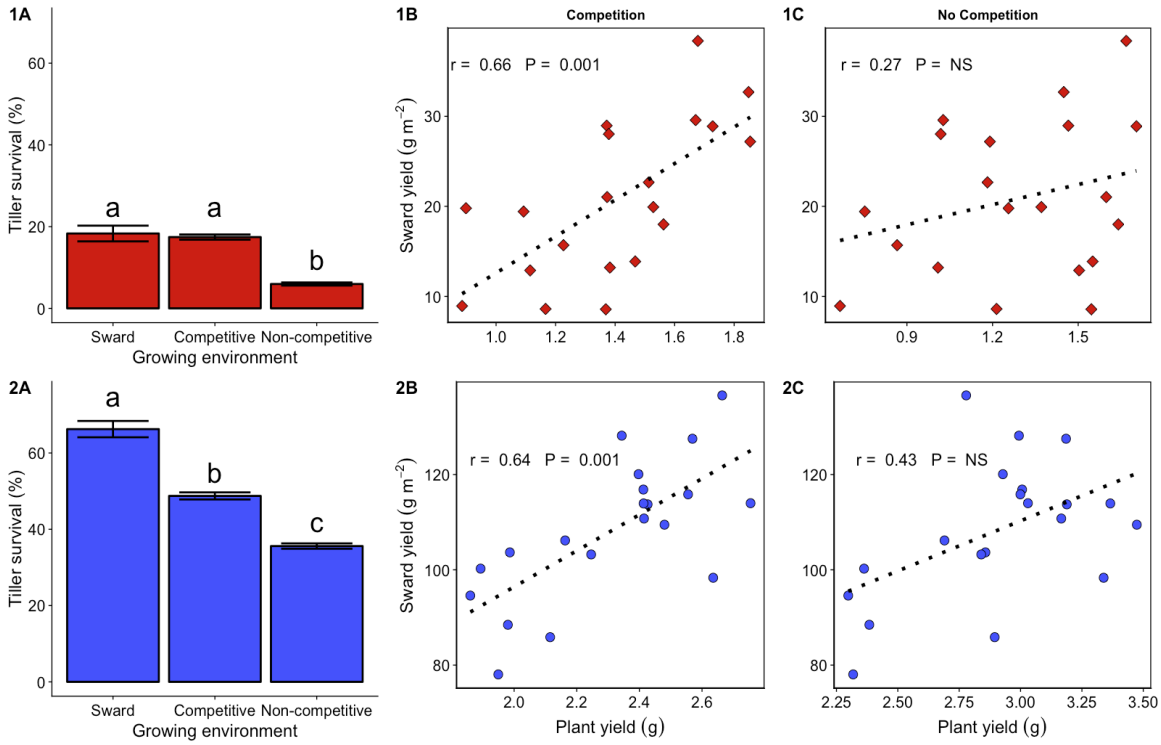


Figure 4.2 - 1A and 2A) Winter survival at Roseau and St. Paul in the sward, competitive and non-competitive environments. 1B and 2B) Relationship between competitive spaced plant yield and seed production plot yield. 1C and 2C) Relationship between non-competitive spaced plant yield and seed production plot yield. Comparisons were made on an entry mean basis with Roseau always represented in red (1) and St. Paul in blue (2). Correlations were calculated using Pearson rank coefficients ($\alpha = 0.05$).

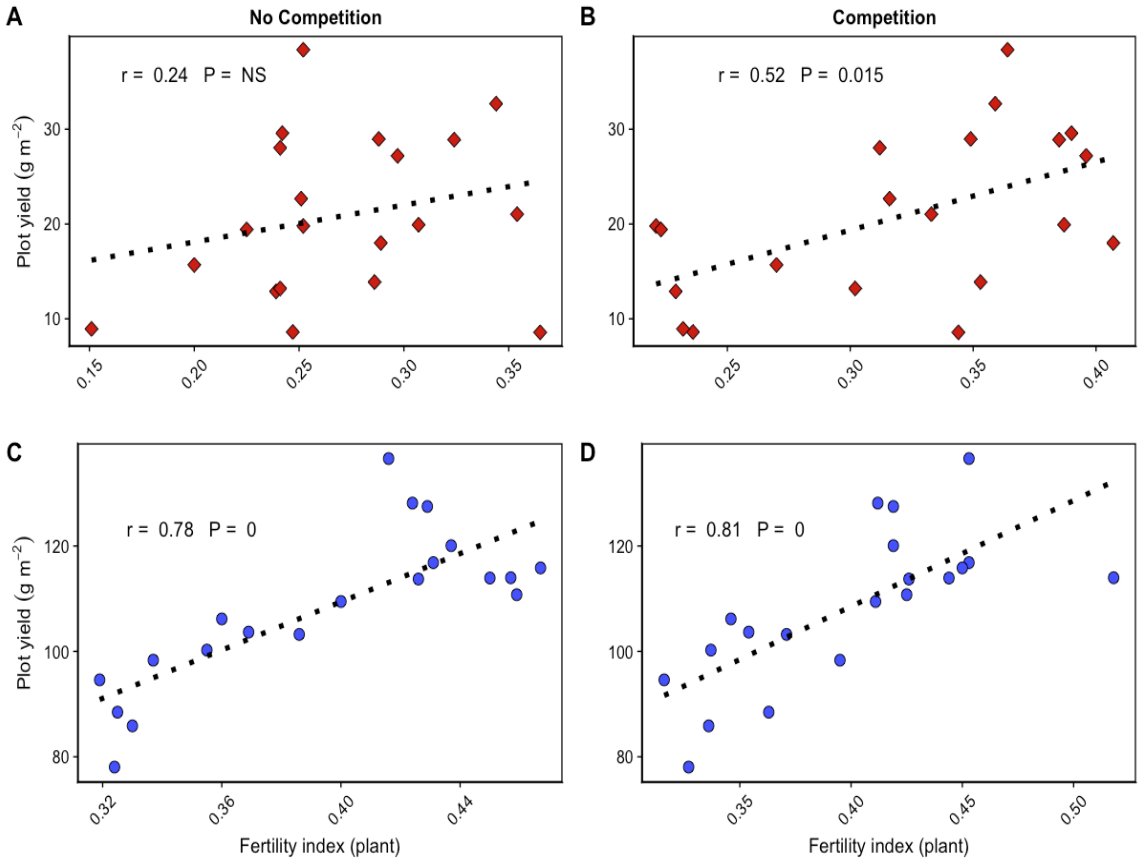


Figure 4.3 - Relationship between spaced plant fertility index and seed yield in the sward environment across two locations, Roseau (red, top) and St. Paul (blue, bottom). A and C) fertility index values calculated on an entry mean basis in a non-competitive spaced plant environment. B and D) fertility index values calculated on an entry mean basis in a competitive spaced plant environment. Fertility index is the proportion of seed to total spike biomass based on measurement of 10 spikes.

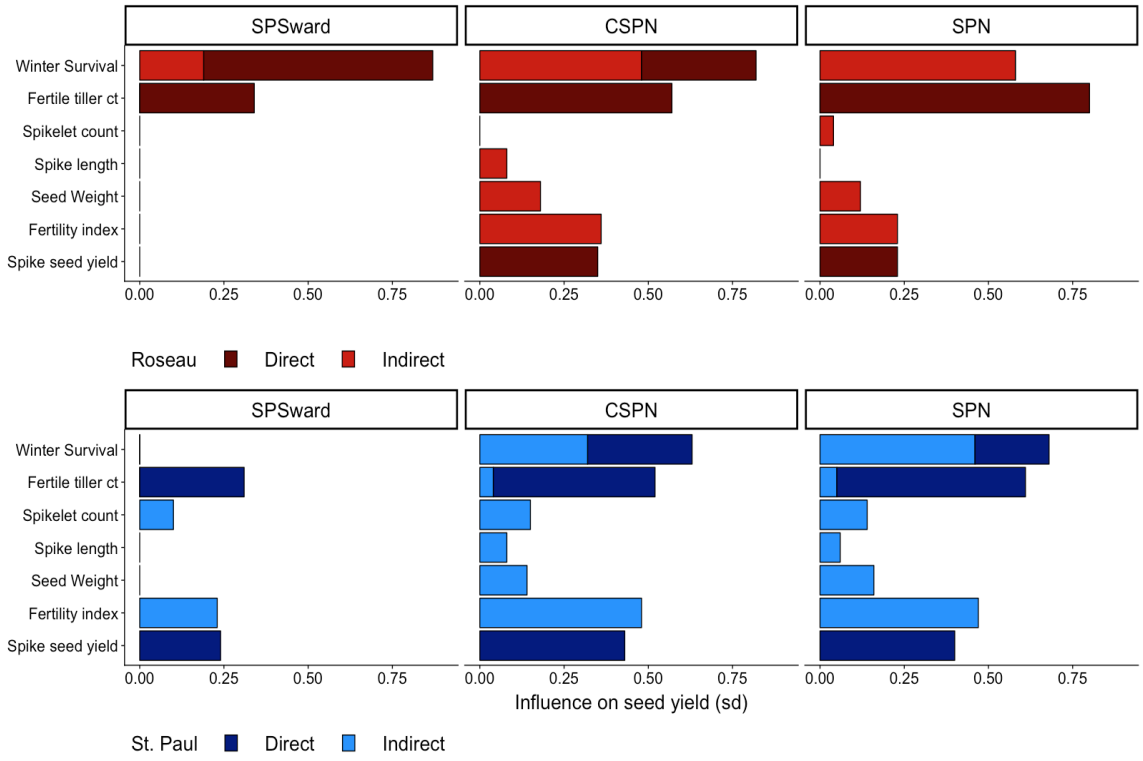


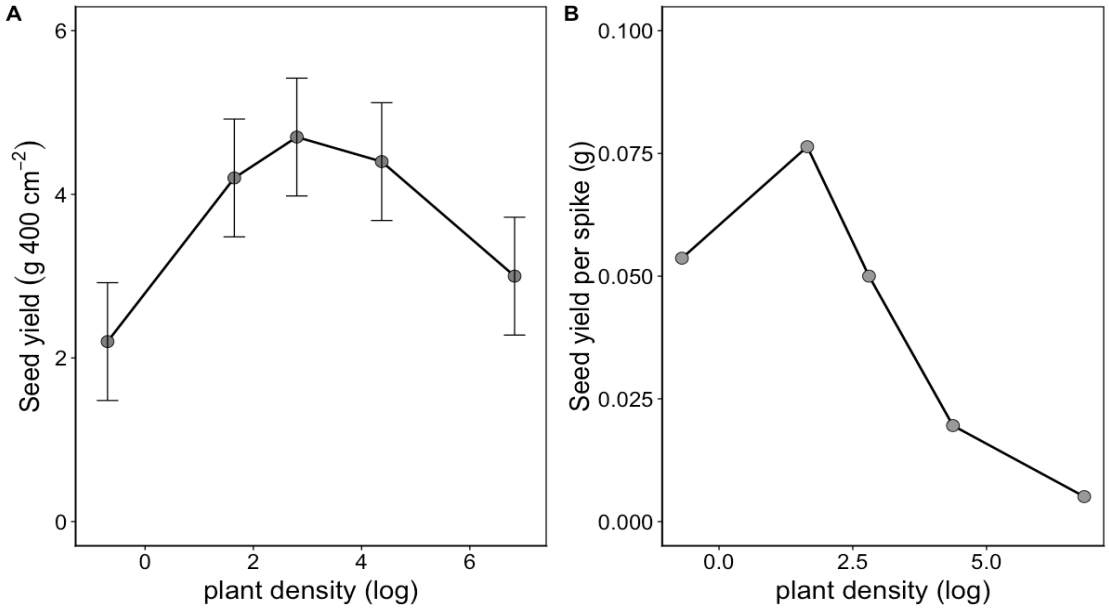
Figure 4.4 - Plotted indirect and direct effects are measured in units of standard deviation and were calculated from six SEM models based on the theoretical seed yield construct (Fig. 4.1). Direct and indirect influences of all yield components are shown for seed production sward (SPSward), competitive spaced plant nursery (CSPN), and spaced plant nursery (SPN) environments. Full SEM diagrams are included in Supplemental Fig. 4.4.

Supplemental Tables

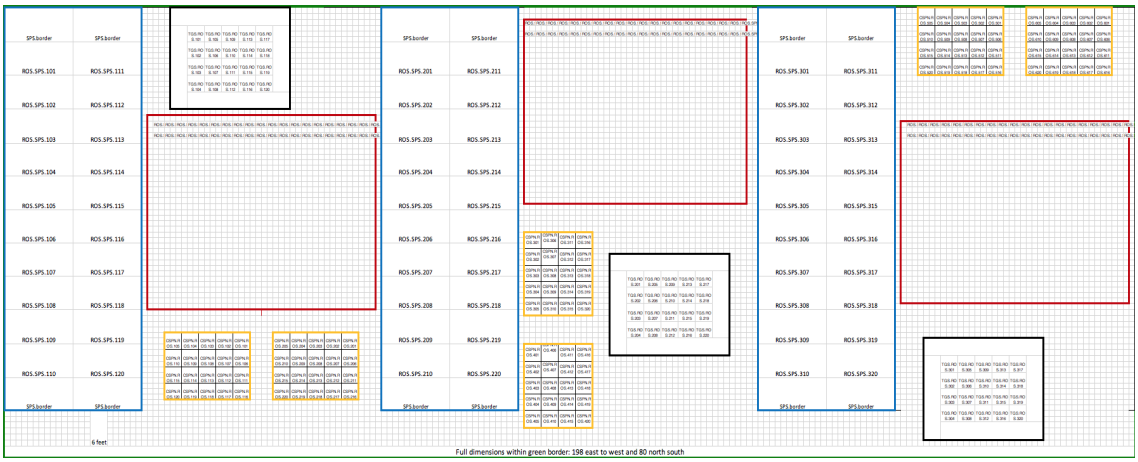
Supplemental Table 4.1 - Experimental entries included in both the turfgrass and seed production trials. Check entries were commercial cultivars and include information on turfgrass quality, seed yield, and winter hardiness. Experimental entries were interrelated half-sib populations contributed by DLF International Seeds breeding program with parental sources originating from commercial cultivars and advanced breeding populations from either Minnesota or Connecticut.

Entry designation	Cultivar/selection name	Germplasm type	Seed source	Parental source	Turfgrass quality rating †	Relative seed yield 2 ‡	Winter hardiness §
PRSM.17.e01	Quebec	Check	DLF	NA	3.1	NA	-2.7
PRSM.17.e02	NK 200	Check	University of MN	NA	NA	Poor	0
PRSM.17.e03	Metolius	Check	Peak Plant Genetics	NA	6.1	NA	NA
PRSM.17.e04	Fiesta_4	Check	DLF	NA	4.7	Moderate	-2.4
PRSM.17.e05	Arctic_Green	Check	University of MN	NA	4.3	Excellent	-0.7
PRSM.17.e06	NGSPRWH-2-12	Experimental	OR	NA	NA	NA	NA
PRSM.17.e07	PR-24-16-06	Experimental	Connecticut	PSG.1037-12K	NA	NA	NA
PRSM.17.e08	PR-24-16-08	Experimental	Connecticut	Zoom	NA	NA	NA
PRSM.17.e09	PR-24-16-09	Experimental	Minnesota	PR.580	NA	NA	NA
PRSM.17.e10	PR-24-16-12	Experimental	Connecticut	Bandalore	NA	NA	NA
PRSM.17.e11	PR-24-16-14	Experimental	Connecticut	PR.562	NA	NA	NA
PRSM.17.e12	PR-24-16-17	Experimental	Minnesota	Karma	NA	NA	NA
PRSM.17.e13	PR-24-16-21	Experimental	Minnesota	PR.524C	NA	NA	NA
PRSM.17.e14	PR-24-16-22	Experimental	Connecticut	Diligent	NA	NA	NA
PRSM.17.e15	PR-24-16-23	Experimental	Minnesota	Monsieur	NA	NA	NA
PRSM.17.e16	PR-24-16-25	Experimental	Minnesota	Aspire	NA	NA	NA
PRSM.17.e17	PR-24-16-26	Experimental	Connecticut	PSG.HLT	NA	NA	NA
PRSM.17.e18	PR-24-16-28	Experimental	Connecticut	PR.583	NA	NA	NA
PRSM.17.e19	PRWHNGS-1-14	Experimental	OR	NA	NA	NA	NA
PRSM.17.e20	WH-PR-11-03	Experimental	OR	NA	NA	NA	NA
† Turfgrass quality determined by national turfgrass trial data.							
‡ Seed production capacity determined by MTSC reports relative to NK200 (2013-2017).							
§ Winter survival as reported by Minnesota Turf Seed Council in relation to NK200 in 2016 (1 unit = 12.5 % change in tiller survival).							

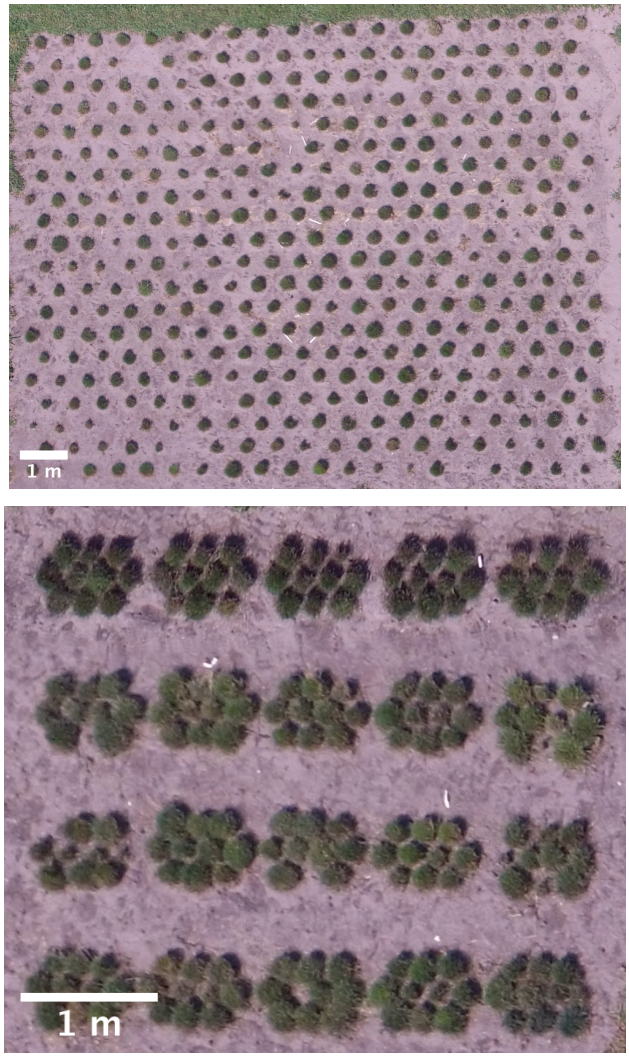
Supplemental Figures



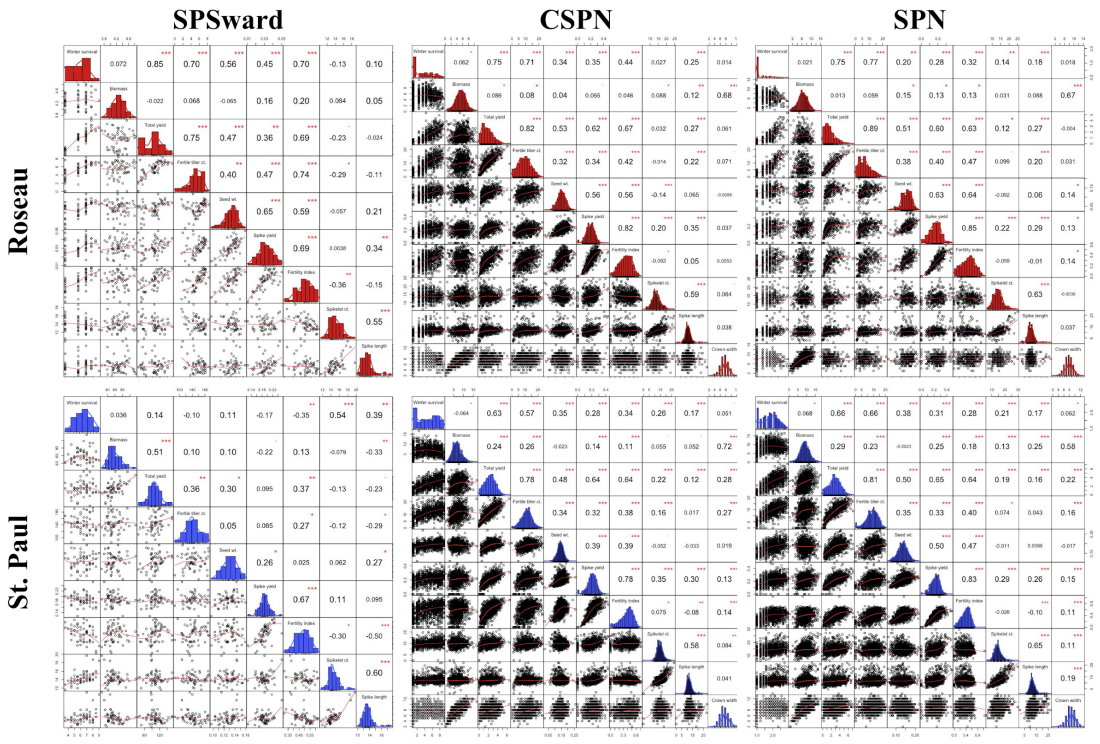
Supplemental Figure 4.1 - Data adapted from Donald (1954) Table 5 on the yield response of Wimmera ryegrass to changes in planting density. Plant densities have been log transformed for ease of observation, exact density levels were 0.5, 5.2, 16.5, 79.2, and 922 plants 400 cm^{-2}. A) seed yield response with error bars representing minimum significant difference (P = 0.05). B) Seed yield per spike in response to different densities, calculated as a proportion from columns [‘Wt. of Seed’/‘Ear-bearing Tillers’] in Table 5 and thus does not have significance reported.



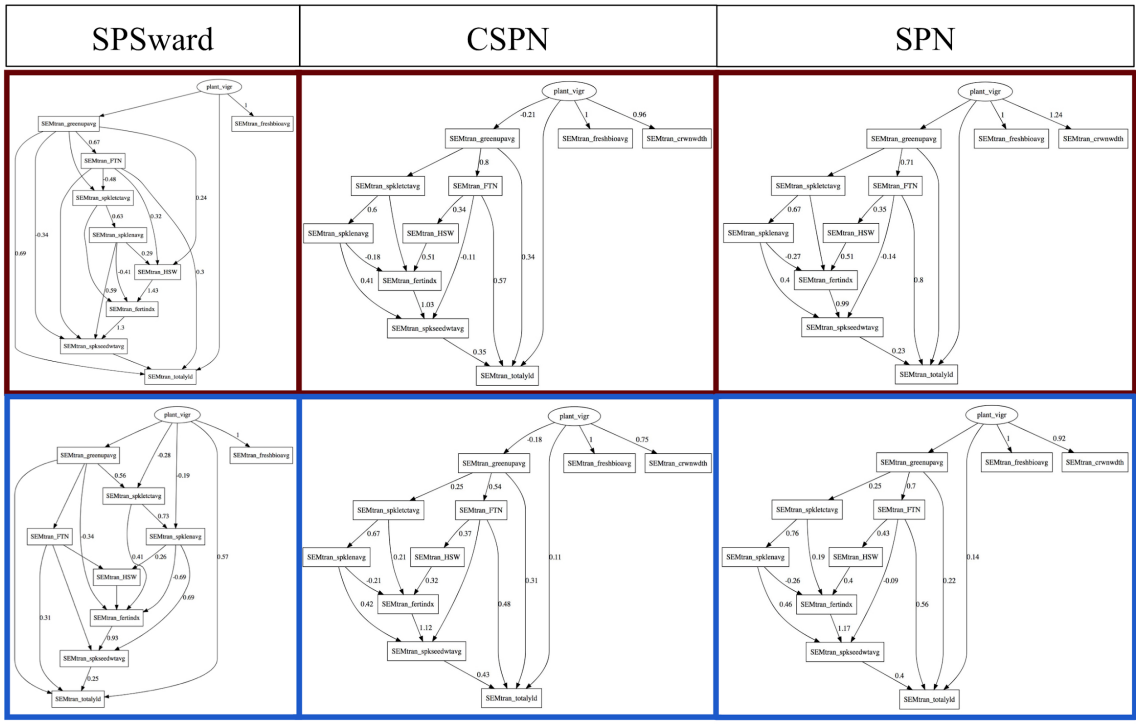
Supplemental Figure 4.2 - Field plan displaying the configuration of the three growing environments: spaced plant nursery (SPN) is outlined in red, competitive spaced plant nursery (CSPN) outlined in orange, and seed production sward (SPSward) outlined in blue. Turfgrass plots are outlined in black but were not included in this study. Each block includes a randomized arrangement of all three growing environments.



Supplemental Figure 4.3 - Images of a single block of both CSPN and SPN growing environments. Standard spaced plant block (top) consisted of 400 plants 0.61 m apart arranged in a 20x20 Latin square. Competitive spaced plant (bottom) included 20 competitive units each made up of 10 plants 0.2 m apart from the same entry.



Supplementary Figure 4.4 - Bivariate spearman rank correlation matrix for seed production sward environment. Bottom panels represent bivariate plot diagrams with smoothed trend lines. Diagonal panels display a histogram showing the distribution on a proportion basis for each trait. Top panels show Spearman rank correlations between variables with '*' $P <$ 0.05, '**' < 0.01, '***' < 0.001.



Supplemental Figure 4.5 - Structural equation model output for seed production environment top) Roseau SPSward, CSPN, and SPN; bottom) St. Paul SPSward, CSPN, and SPN. Four additional paths were added to each SPSward model. These had only very minor implications on the direct and indirect paths overall.

Chapter 5

Predictive ability of perennial ryegrass spaced plants for turf and seed production growing environments

Introduction

Turf type perennial ryegrass (*Lolium perenne* L.) breeding programs have been actively selecting for turfgrass quality and seed yield capacity since the mid 1950s. Gains in cultivar performance over the last 30 years in particular have been recognized for turfgrass quality, density, and crown rust resistance. However, many traits such as winter survival, density/tillering, vegetative biomass and seed yield have not been substantially improved (Sampoux et al., 2013; Casler et al., 1996). New breeding methodology could improve both gains from phenotypic selection and also allow for more accurate marker trait associations by improving predictive ability across environments.

Cool-season turfgrass breeding programs utilize spaced plants as the initial unit of selection for turfgrass quality and related traits such as color, density, and texture. Traits related to seed production may also be targeted at this stage, including vegetative biomass, heading date, height, and seed yield. Spaced plants are a convenient way to observe plants and maintain the identity of individuals (Casler and Van Santen, 2010). Standard spaced plant environments typically induce little to no competition with densities ranging between 2-4 plants m^{-2} (Foster, 1971; Elgersma, 1990; Waldron et al., 1998; Casler and van Santen, 2008). This low density allows for selection to occur both within and between families, which can lead to increases in selection efficiency. Moreover, seed stock is a critical consideration early in the breeding pipeline as half- or full-sib families typically do not have enough seed for even a small turfgrass or seed sward. This also necessitates a spaced plant design. Progress from selection in spaced plantings can be particularly successful for some high heritability traits such as heading date and some diseases (Casler, 2010). Unfortunately, there is substantial evidence that performance under noncompetitive environments is not indicative of a sward (Lazenby and Rogers, 1964; Hayward and Vivero, 1984; Bugge, 1987; Stratton and Ohm, 1989; Elgersma, 1990; Elgersma et al., 1994; Waldron et al., 2008; Sykes et al., 2016).

Genotype-by-plant spacing interaction studies are particularly common in forage breeding and have been conducted for species such as perennial ryegrass (Lazenby and Rogers), tall fescue (*Festuca arundinacea* Schreb.; Waldron et al., 2008), switchgrass (*Panicum virgatum* L.; Sykes et al., 2016), and orchardgrass (*Dactylus glomerata* L.; Stratton and Ohm, 1989). Similar studies have also focused on predictive ability of seed

yield in perennial ryegrass (Bugge, 1987; Elgersma et al., 1994). Although some authors report high predictive ability for complex traits, most conclude that wide spacing in selection nurseries inflate phenotypic differences and can lead to low correlation with sward environments. This is likely derived from phenotypic response to inter and intra plant competition, which has been observed in both tillering and seed yield across plant densities from 12 to 22,500 plants m^2 in *Lolium* (Donald, 1954). This change in plant density would be equivalent to spaced plants under moderate competition to a turfgrass sward and is the reason for poor phenotypic reproducibility (Elgersma, 1990). For instance, gains from selection for forage yield in spaced plants and sward have been shown to be greatly inflated in cereal rye (*Cereal secale* L.) (Bruckner et al., 1991).

Knowledge of poor space-plant predictive ability in cool-season grasses has been reported for the same length of time that turf-type perennial ryegrass programs have existed (Ahlgren et al., 1945). Nevertheless, the convenience and non-species-specific application of wide spacing has allowed it to persist. Sedcole and Clements (1973) tested a competitive design with 100 plants m^{-2} and argued against decreasing the spacing between plants, stating that increasing plant density made genotypes difficult to observe and that gains from selection could be made under noncompetitive conditions. This same observation was made by Lazenby and Rogers (1964) with regards to plants grown at 175 plants m^{-2}, however they noted that perennial ryegrass plants at a density of 19 plants m^{-2} behaved similarly to those in the closer spacing. This lower density allowed for more accurate phenotypic data to be collected with respect to a sward environment (Supplemental Fig. 1).

Many different selection environment strategies have been developed to improve prediction accuracy and reduce seed requirements. Hayward and Vivero (1984) compared both microswards (greenhouse flats broadcast seeded at 210 plants m^{-2}) and imitative drill (narrow spaced plantings at 17 plants m^{-2} within rows with wide spacing between rows) with a standard spaced design of 3 plants m^{-2} and showed that both competitive designs improved prediction for biomass in a sward. Foster (1973) also tested perennial ryegrass populations in microswards (900 plant m^{-2}) but did not find it to be superior to spaced plants for biomass yield.

Other authors have suggested planting a competitive species around perennial ryegrass spaced plants such as (*Poa pratensis* L.) or timothy (*Phleum pratense* L.) as

competitive surrogates (Humphreys, 2010). Samuel et al. (1970), who found that spaced plants (2 plants m^{-2}) were negatively correlated with sward yield, attempted to solve the problem of rank discrepancies between spaced plants and swards for biomass through linear regression to determine a common conversion factor; this, however, had very dubious results that are not contemporarily used. Bhandari et al. (2013) utilized a modified selection strategy with standard spaced plantings for switchgrass that involved a honeycomb design in which encircling plants were used to estimate the yield of each plant. Synthetics resulting from each method revealed that greater gains were obtained from sward selection, but authors concluded that the spaced plants were more efficient with regards to space and time to make selections. These aforementioned modifications to experimental design or analysis technique may improve accuracy, but do not address the underlying issue of genotype by environment interaction.

While the studies reviewed above give strong evidence that standard spaced plant densities are ill-suited for many quantitative traits, increasing plant densities to more than 100 plants m^{-2} needs be avoided due to concerns with overcrowding (Table 5.1). The size of a standard perennial ryegrass crown must be considered to effectively design a nursery that not only fosters competition but also allows for ease of observation. Waldron et al. (1998) developed a topcross population to breed for increased winter hardiness in turf type perennial ryegrass in Minnesota and found that crown widths varied considerably but had an average width of 19.1 cm. This indicates that spacing should not drop below 19.1 cm (~27 plant m^{-2}) to avoid loss of plant identity and increase efficiency of phenotyping.

Phenotyping through image analysis has become increasingly popular over the last two decades throughout turfgrass and forage research (Patrignani and Ochsner, 2015; Yu et al., 2019). Simple RGB images allow for not only the long-term storage of two-dimensional phenotypes, but also a means to analyze data in the off-season and thereby collect more data. Digital image analysis, although not sufficient to measure some traits, has pragmatic potential to replace ordinal visual estimates with quantitative measures. Rust severity and turfgrass density are common targets for image analysis in turfgrass, however winter survival may also be amenable to assessment using image analysis (Díaz-Lago et al., 2003; Karcher and Richardson, 2005).

The objective of this study was to test two different spaced plantings based on both the standard spaced-plant density of 3 plants m^{-2} (SPN) and a hypothesized optimal competitive spaced plant density of 23 plants m^{-2} (CSPN). Twenty perennial ryegrass entries were used to determine whether these spaced plant growing environments were able to predict seed production (SPSward) and turfgrass sward (TGSward) environments.

Materials and Methods

Plant material

Perennial ryegrass experimental entries consisted of 15 interrelated populations and 5 cultivars. Interrelated populations were aggregated from several half-sib families all originating from the same source. The maternal sources for each population were selected from advanced synthetic material in the DLF Pickseed - International Seeds breeding program that was tested at the University of Minnesota and the University of Connecticut (Supplemental Table 5.2). Parental material was sent to Oregon to be polycrossed. One forage-type and four turf-type and cultivars were chosen based on a wide performance range for traits including winter survival, turfgrass quality, and seed yield capacity (Supplemental Table 5.1).

Experimental locations and management

Seed production and turfgrass swards were seeded and spaced plant nurseries were transplanted in June 2017 (Fig. 5.1). The experiment was tested at two locations in Minnesota, one at the Minnesota Agricultural Experiment Station at the University of Minnesota in St. Paul, MN and the other at the University of Minnesota Magnusson Research Farm in Roseau, MN. The two locations were located approximately 560 km apart. Roseau County is the major seed production region in Minnesota and the University of Minnesota turfgrass and seed breeding program is located in St. Paul. The soil at Roseau is a Borup silt loam soil with 2.7% organic matter and a pH of 7.7, whereas St. Paul is a Waukegan silt loam soil with 4.8% organic matter and a pH of 6.4.

Sward and spaced plant growing environments were maintained with the same management practices. Upon establishment, plots and nurseries were fertilized using a Gandy drop spreader (Gandy Company, Owatonna, MN) at a rate of 24.4 kg N ha^{-1}, 40.1

kg P ha^{-1}, and 24.4 kg K ha^{-1}. In late September 2017, plots were fertilized using a Gandy drop spreader at a rate of 0.0 kg N ha^{-1}, 33.6 kg P ha^{-1}, and 33.6 kg K ha^{-1}. Broadleaf weeds were controlled in summer of the establishment year (July 6th and June 19th at St. Paul and Roseau, respectively) with a single application of bromoxynil and MCPA [(3,5-dibromo-4-hydroxybenzonitrile and 2-ethylhexyl ester of 2-methyl-chlorophenoxyacetic acid); Bison, BASF, Research Park Triangle, NC] at a rate of 0.28 kg a.i. ha^{-1}. Fall and early spring weed seedlings were controlled with a single application of pendimethalin (N-(1-ethylpropyl)-3,4-dimethyl-2,6-dinitrobenzenamine; Prowl H_2O, BASF, Research Park Triangle, NC) at a rate of 2.12 kg a.i. ha^{-1}. Fall broadleaf weeds were controlled (Sept 25th and Oct 4th at St. Paul and Roseau, respectively) with an application of 2,4-D (2, 4-dichlorophenoxyacetic acid; Shredder, Winfield St Paul, MN) at a rate of 0.40 kg a.i. ha^{-1}. In the seed production year, fertilizer was applied in early May at a rate of 123.1 kg N ha^{-1}. Broadleaf weeds were controlled in late May with an application of 2,4-D (2, 4-dichlorophenoxyacetic acid; Shredder, Winfield, St. Paul, MN) at a rate of 0.27 kg a.i. ha^{-1} and dicamba (3,6-dichloro-o-anisic acid; Sterling Blue, Winfield, St. Paul, MN) at a rate of 0.27 kg a.i. ha^{-1}. All chemical applications were applied using a CO_2-powered bicycle sprayer equipped with a 2.75 m boom and eight 1002 TurboTeeJet® (TeeJet, Springfield, IL) nozzles operating at 186 kPa in water equivalent to 117 L ha^{-1}.

In both years, climate data for Roseau were collected from the North Dakota Agricultural Weather Network, Fox station (48.878013, -95.847207) and a weather station at the University of Minnesota St. Paul campus (44.995031, -93.185839). Growing degree days (GDD) were calculated using a base temperature of 0°C. Data were taken on soil temperature 10 cm below the soil surface.

Experimental environments

Seed production swards were seeded with a Hege 1000 small-plot seeder (Hege Equip. Inc; Colwich, KS). Each plot was 3.6 m long with 10 drilled rows 15 cm apart. Each plot was seeded with 6.5 kg ha^{-1} of pure live seed resulting in approximately 75 seeds per linear meter of row (380 plants m^{-2}). Each location had three blocks, along with seeded border plots, arranged in a randomized complete block design. Turfgrass plots were 0.91 x

0.91 m. Seeding was done by hand at a rate of 90 kg ha^{-1} pure live seed and then raked-in resulting in approximately 0.5 seeds cm^{-2} (5000 plants m^{-2}).

Plant material for SPN and CSPN were seeded in April and allowed to grow in the greenhouse until transplanted into the field during the first week of June. Standard non-competitive spaced plant nurseries were arranged in a 20 x 20 Latin square with three squares (blocks). Plants from each entry were randomized within each row and column and were planted 0.61 m apart (3 plants m^{-2}), which is a typical distance that minimizes competition. Competitive spaced plants were transplanted into units that facilitated moderate competition and consisted of 10 plants of the same entry. Competitive units were aggregated so that each block contained 200 plants (20 entries x 10 plants). The CSPN environment included six blocks instead of three to minimize localized winterkill (Fig. 1), however the same number of plants was included in the SPN nursery. Plants in CSPN were planted 0.21 m apart (23 plants m^{-2}), which fostered competitive interactions among plants, but allowed observation on spaced plants (Lazenby and Rogers, 1964). Perennial ryegrass breeding populations are very diverse even within families. To make accurate selections a breeder must then phenotype large numbers of individuals (>20) to make meaningful selections within any population or calculate genetic components (Bolaric et al., 2005); thus, both CSPN and SPN environments were represented by 1200 plants at each location (Fig. 2).

Seed production, TGSward, CSPN, and SPN environments were arranged in a randomized complete block design within location (Fig. 5.1). This was accomplished by randomizing growing environments at the block level. Blocks were arranged in this way so that statistical comparisons could be made between entries across growing environments.

Data Collection

Seed production establishment was measured by counting seedlings at both locations 14 d post-seeding on two 0.5 m transects of row. Every two weeks relative chlorophyll index was measured using a FieldScout CM 1000 (Spectrum Technologies, Aurora IL) and imaging was conducted (see image analysis). Crown rust severity was rated visually using the Modified Cobb scale (Roelfs, 1992). Fresh biomass was collected in the

field in late August 2017 by harvesting three 0.5 m transects at a cutting height of 7 cm from each plot. Winter survival was measured as the mean of repeated visual estimates of winter survival on a 1-9 scale with a 5 equal to 50% green tillers within a row (Hulke et al., 2007). Growth stage assessments were conducted in the seed production year three times at St. Paul and twice at Roseau. The plant maturity scale included eight categories and was adapted from Gustavsson (2011) with: 11 = developing vegetative leaf with little elongation, 31 = noticeable stem elongation, 39 = flag leaf collar just visible, 45 = inflorescence was palpable the flag leaf sheath, 50 = spikelet being visible above the collar of the flag leaf, 59 = inflorescence bearing internode being visible above the collar of the flag leaf, 61 = the first anther being visible, 69 = all spikelets were finished flowering but some dried anthers were still visible. The median tiller growth stage was visually assessed on individual spaced plants, with nine spike sample recorded along a 2.0 m transect. Height was measured post anthesis with three samples taken per plot as the length of fertile tiller from base of plant to top of the spike on approximately 10 culms. Lodging was measured through repeated visual estimates of stem angle on a 1-7 scale where 1 = 0º, 4 = 45º, 7 = 90º degree angles. Culm width was determined by taking the mean width of four stems approximately 10 cm below the seed head. Fertile tiller number (FTN) samples were taken by cutting four-15 cm transects along a row within each plot so that all culms were exposed, and a flush surface was created, followed by a second cut that was made so that a 3-4 mm segment from each culm dropped into a bin. These segments were funneled into an envelope, cleaned using a General Seed Blower (Seedburo Equipment Company, Des Plaines, IL) and counted using a Seed Count S-JR seed counter (Data Technologies, Kibbutz Tzora, Israel). Whole plot yield was measured by collecting two 1 m^2 swaths of perennial ryegrass per experimental unit at approximately 400 g kg^{-1} moisture content. All samples were dried for 5 d at 35 °C in a forced air dryer and processed in a LD350 thresher with a 2.5 x 7.5 mm concave and processed at 800 rpm (Wintersteiger Corp., Ried im Innkreis, Austria). Following threshing, seed and chaff were separated by passing the processed seed through a 1.625 x 9.525 mm sieve. Light seed was removed by aspiration using a Superior® Fractionating Aspirator (Carter-Day Int. Inc., Minneapolis, MN).

Turfgrass sward establishment was assessed on a 1-9 visual scale with a 5 equal to 50% of the seedlings filling in a 15 x 15 cm square grid. Turfgrass quality, genetic color,

texture, density, rust incidence, height, and image data were collected every two weeks in both years. All turfgrass quality traits were assessed according to the National Turfgrass Evaluation Program guidelines (Morris and Shearman, 2006). Image data were collected every two weeks at both locations three weeks after seeding (see image analysis). Custom diagrammatic scales were used for density and genetic color (1-9, 9 being most favorable) to maintain consistent rating across locations (Supplemental Fig. 5.2). Turf height was assessed before each mowing by measuring three points in each plot using a TurfChek II grass height gauge (Turf-Tec International, Tallahassee, FL). Winter survival was measured as the mean of repeated visual estimates of spring green-up on a 1-9 scale with a 5 equal to 50% green tillers within a plot. Stemminess was measured visually (Islam et al., 2013), but was also quantified through visual stem counts and the proportion of stems detected in images.

Spaced plants were measured similarly for both SPN and CSPN growing environments. Imaging was conducted twice in the establishment year and once in the production year. In the establishment year, data was collected once on turfgrass quality, genetic color, texture, and crown rust infection using the same methods that were used in the TGSward environment. Plant height and biomass of each plant were collected as fresh weight on the same days SPSward was collected. Winter survival was measured both as a binomial (1 = alive and 0 = dead) and on a 1-9 scale with a 5 equal to 50% green tillers within a plant. Growth staging, culm width, FTN, and height were measured similarly to the SPSward environment. Seed yield of whole plants was measured on an individual plant basis with identity of each plant maintained through a barcoding system. Drying, threshing, and seed cleaning was conducted with the same method as plot samples.

Image analysis technique

All images were taken with a Nikon D300 digital SLR camera and saved as high-quality JPEG files of 4288 x 2848 and 3216×2136 pixels for sward and spaced plants, respectively. Imaging was conducted under ambient conditions in the establishment year between 1200 and 1400 hours with a fixed monopod mounted 80 cm above the ground for turf plots and spaced plants and 160 cm for seed plots, squarely above the plot or plant.

Camera settings were shutter speed of 1/160 s, aperture setting of F8, white balance 5000K, and a focal length of 80 mm.

In the production year, images were taken using the same camera but with a light box constructed of an aluminum frame (0.8 m wide x 0.6 m long x 0.9 m tall) with coroplast walls lined with mylar to eliminate light infiltration. The interior was lit by 200 cm of high color rendering index LED light strip with an output of 221 lumens 0.3 m^{-1} and a 4000K color value (superbrighLED.com). The lights were powered by a rechargeable Talentcell 12-volt lithium ion battery pack (talentcell.com). Camera settings for light box images were shutter speed 1/80 s, aperture setting F2.7, white balance 5000K, and a focal length of 80 mm.

Images were used to quantify lateral growth, crown rust, and winter survival in each environment. Details of minor modifications for each of these traits are described herein; a full description of the accuracy and utility of the analysis technique is described in chapter 1 of this thesis by Heineck et al. (2019). Image analysis was done using ImageJ and R (Schneider et al., 2012; Version 3.6.0; R Core Team, 2018). Core packages in R were 'EBImage' and 'randomForest' (Liaw and Wiener, 2002; Pau et al., 2010). Separate training data sets were made for each year and environment using images from each rating date, with about 300 data points collected from each image to train the model. Individual random forest models were generated for each year and environment. Image analysis output at each time point was inspected visually for any abnormalities in the processing pipeline. Crown rust pustules were separated from noise with adaptive thresholding using a disc shape of 9 pixels in diameter. Winter survival was quantified as the proportion of living to dead plant tissue to account for both inter-row space and plant size differences. Winter killed plant tissue was difficult to quantify due to the noise produced by decomposing plant tissue and algae growth (Fig. 3). Both fillhull() and adaptive thresholding with a large window size for both Roseau and St. Paul (51-251 respectively). Results of this image analysis can be seen in Supplementary Fig. 5.

Statistical analysis

Analysis of all data was conducted in the R environment (R Core Team, 2018). Analysis of variance and analysis of covariance was conducted using a fixed effects model

in the 'lme4' package (Bates et al., 2015). When locations were analyzed together, location x block interaction was included a random effect. Entry marginal means were estimated and separations were conducted using the 'emmeans' package and a Tukey HSD test to detect significant differences ($\alpha = 0.05$) (Lenth, 2018). Bivariate relationships were explored through simple linear regression with large numbers of regressions conducted using the 'PerformanceAnalytics' package. Correlation coefficients were based on Spearman rank correlation calculations. When regression was conducted between growing environments (e.g. seed yield collected from SPN vs. SPSward) entry means were used; however, when regression was done within environment data points from each block were used.

Results and Discussion

This study explored the potential predictive ability of spaced plants in relation to both turfgrass and seed production swards. The noncompetitive SPN design minimized inter-plant competition, while the CSPN design fostered moderate competition among related individuals. Phenotyping was conducted on 20 turf-type perennial ryegrass entries using both traditional methods as well as a new image analysis technique (Table 5.1 and Fig. 5.3). The experiment was conducted at two locations in Minnesota including Roseau, which is the central seed production region located in the far northern part of the state, and St. Paul where the University of Minnesota turfgrass and forage breeding programs are located. Spaced plants and swards were measured over two growing seasons: the establishment year in the fall and the seed production the following summer.

Image analysis performance

Image analysis was used to quantify several traits that have been traditionally assessed through visual ratings in both spaced plant and sward environments (Fig. 5.3). Each method was checked against manually collected data to estimate accuracy on an entry-within-block-mean basis. The ability of green pixels to approximate the biomass of a spaced plant was tested via two methods: 1) number of green pixels at the end of the growing season; and 2) daily increase of green pixels over 28 d (pixel d^{-1}). Method 1 yielded sporadic success, and was only accurate in the CSPN Roseau environment (r_p =

0.84, P < 0.001), however total green pixels for the SPN was a significant predictor of biomass at both Roseau (r_p = 0.47, P < 0.001) and St. Paul (r_p = 0.75, P < 0.001). Method 2 had slightly better results, albeit not consistent: the accumulation of pixels per day was a significant predictor of biomass for CSPN at both Roseau (r_p = 0.83, P < 0.001) and St. Paul (r_p = 0.19, P = 0.04) and for SPN at Roseau (r_p = 0.43, P < 0.001) and St. Paul (r_p = 0.81, P < 0.001). The lack of correlation in the CSPN environment is likely due to the difficulty in accounting for leaf area index as the neighboring plant canopies overlap (Patrignani and Ochsner, 2015). This relationship was further investigated when green pixel number was correlated with biomass in SPSward environment and unreliable coefficients were yet again obtained for both Roseau and St. Paul (r = 0.53 and 0.51, P < 0.001 respectively).

Crown rust severity was only measured visually in the establishing year and only when sufficiently high infection levels occurred for visual detection in St. Paul; however, image analysis was able to detect low severities of crown rust in Roseau. Correlation coefficients between visual and computer ratings were adequate in both CSPN (r_p = 0.82, P < 0.001) and SPN (r_p = 0.79, P < 0.001) nursery designs. Correlation coefficients between visual and computer ratings were good in both TGSward (r_p = 0.79, P < 0.001) and SPSward (r_p = 0.92, P < 0.001) nursery designs. Winter survival is typically estimated visually by approximating the proportion of green tillers to dead tillers (Hulke et al., 2007). Image analysis was highly adept at quantifying winter survival and was likely more accurate than visual ratings (Fig. 5.3). Competitive nursery predictions were relatable to visual ratings at both Roseau (r_p = 0.92, P < 0.001) and St. Paul (r_p = 0.96, P < 0.001). Standard nursery predictions were just slightly lower at Roseau (r_p = 0.91, P < 0.001) and St. Paul (r_p = 0.89, P < 0.001). Stemminess has been visually estimated, but the proportion of stems is difficult to determine (Fig. 3; Islam et al., 2013). Stems in the TGSward environment were manually counted within the image and compared to computer generated proportions, which yielded a high correlation (r_p = 0.88, P < 0.001). Because of the strong relationship between computer and visual measures, crown rust severity, winter survival, and stemminess were analyzed using image acquired data. Utilizing image analysis holds several benefits over visual estimates including lower employee skill level requirement and

the permanent recording of digital phenotypes, which can be analyzed at any time throughout the year.

TGSward and SPSward trait performance

The objective of this study was to determine if spaced plants can predict sward traits; however, successful prediction matters little if entries do not differ significantly for traits of interest. This would inevitably lead to false conclusions either because of a lack of variability to predict or a lack of broad sense heritability. Furthermore, the presence of trait instability across locations is essential knowledge if materials are bred for multiple environments.

Establishment was analyzed for both sward environments to determine if differences occurred between locations and or entries. Analysis of variance found that TGSward establishment was fairly consistent across locations ($F_{1,4} = 4.6$, $P = 0.10$) and entries ($F_{19,76} = 1.6$, $P = 0.09$) with a nonsignificant interaction term. Entries in the SPSward environment did not differ in establishment ($F_{19,76} = 1.5$, $P = 0.09$), but St. Paul had significantly larger plant populations than Roseau (49 vs. 32 plants m^{-1}, respectively). These results showed that entries were successfully established and that trait differences were not due to poor germination. Turfgrass quality ratings were averaged over the establishment and production years. A significant entry-by-location interaction was found for turfgrass quality in the establishment year ($P < 0.001$). Turfgrass quality ranged from 4.5 to 7.3 at Roseau and 2.5 to 6.8 at St. Paul. The range of quality ratings reported here is fairly typical and is similar or greater than that recorded by Johnston et al. (2011) who studied turfgrass and seed yield potential of Kentucky bluegrass.

Analysis of variance found that average density in TGSward differed between entries ($P < 0.001$), but no effect of location or interaction was observed ($P > 0.10$). Similar results were found in the SPSward environment with respect to biomass where entries were significantly different ($P < 0.001$) but did not change rank between locations. Turfgrass and SPSward environments were both heavily infected with crown rust (*Puccinia coronata* f. sp. *lolii*) in the establishing year, but not in the production year. Stem rust (*Puccinia graminis* ssp. *graminicola*) was not present in the SPSward environment in either year. Image analysis was used to quantify crown rust severity at both locations. Entry

significantly interacted with location in the TGSward for crown rust severity ($P < 0.001$). Turfgrass entry means ranged from 0.1 to 12.0 % severity at Roseau and from 0.1% to 1.3% severity at St. Paul. Entries in the SPSward environment also differed for crown rust severity ($P < 0.001$), but rankings were consistent across location ($P > 0.10$) with mean severities ranging from 0.0 to 5.7%. These severity values seem fairly small; however, it is important to remember that severity cannot exceed 37% on any given leaf.

Winter conditions were substantially different between the two locations with soil temperatures below freezing at Roseau and St. Paul for 155 and 110 d, respectively. Additionally, in the production year there was very little rainfall at Roseau compared to St. Paul (18 vs. 37 cm, respectively). These conditions led to winter kill at both locations, but especially high levels of tiller mortality were observed at Roseau. Analysis of variance found no differences among entries in the TGSward environment for winter survival, and although locations were not significantly different ($P = 0.12$) there was a large difference in tiller survival between Roseau and St. Paul (31 vs 52% respectively). In the SPSward environment, entry ranking for winter survival was only moderately consistent across locations ($P = 0.06$) and varied greatly at Roseau (1 to 16% survival) and at St. Paul (29 to 74% survival).

Turfgrass quality in the second year was highly correlated with winter survival on a plot level ($r_s = 0.81$, $P < 0.001$). Analysis of variance found that mean turfgrass quality differed for both entries ($P < 0.001$) and locations ($P = 0.02$). Cumulative vertical turfgrass growth and stemminess were both measured in the production year as they can negatively influence turfgrass quality in the TGSward environment. Analysis of variance found no entry by location interaction for either trait. However, entries had different vertical growth rates ($P < 0.001$) as well as proportion of stems to green leaves ($P < 0.001$).

Seed yield was highly dependent on winter survival if extreme tiller mortality was observed ($r_s = 0.85$, $P < 0.001$). Repeated measures of growth stage indicated that heading date was much later at Roseau compared to St. Paul. Analysis of variance found that entry rankings were not stable across locations for both stem height and lodging. Entry means for seed yield were also highly variable across the two locations ($P = 0.007$) with Roseau yielding 8.6 to 38.4 g m^{-2} and St. Paul yielding 78.0 to 136.6 g m^{-2}. Due to the variable and extreme impact of winter kill and the generally large location by entry interaction terms,

locations were analyzed separately to determine the predictive ability of spaced plants. Roseau is in Minnesota's grass production region and was an important location to test; however, winterkill and highly variable conditions are not unexpected and often lead to individual location-year analyses (Koeritz et al., 2013; Koeritz et al., 2015).

CSPN vs. SPN trait performance

Increasing competition between plants will likely alter plant phenotype and may lead to more accurate entry rankings for sward performance. Several authors have found that higher levels of phenotypic variance were found in noncompetitive spaced plant versus a sward environment. This wider variance, although seen by some as an advantage in breeding (Sedcole and Clements, 1973), can certainly be problematic when interactions exist between genotype and plant spacing (Supplemental Fig. 1; Lazenby and Rogers, 1964). Furthermore, wider trait variance in spaced plants can lead to inflated predicted sward-performance gains from selection (Bruckner et al., 1991).

In Roseau, turfgrass quality and genetic color were very similar between nursery designs with respect to mean and standard deviation (Table 5.2). Image analysis detected far lower variability for crown rust severity in the SPN vs CSPN (0-0.08% vs 0-0.18%, respectively). More biomass was produced in SPN versus CSPN plants (Table 5.3). Crown width between CSPN and SPN varied little in the establishing year (7.7 vs. 7.9 cm, respectively). A much greater difference in crown width was observed in the production year (12.5 vs. 15.4 cm, respectively), which may be evidence of limited competition in the establishing year. Winter kill led to much lower tiller survival in the SPN compared to the CSPN design (5 vs 22% respectively). This was likely due to snow catch between plants and was also seen in the sward environments. Plant height, lodging, and fertile tiller number in the production year was very similar between CSPN and SPN, however CSPN plants in general had slightly more, taller, and upright fertile tillers (Tables 5.2 and 5.3). Advances through growth stage were similar between spaced plant environments, but seed yield was slightly higher in the CSPN environment (2.81 vs. 2.72 respectively). However, the SPN had a wider standard deviation (Table 5.3).

Trends at St. Paul were very similar to those found in Roseau for turfgrass quality, crown rust, genetic color (Table 5.2). However, the disparity between crown rust severity

in CSPN and SPN was even more drastic (0-0.22 vs. 0-0.72% respectively). Similar to Roseau, SPN had on average larger crowns, but a larger difference was seen at St. Paul in the establishing year. Winter kill led to lower tiller survival in SPN vs CSPN, but not the degree observed in Roseau (Table 5.2). Far more fertile tillers were produced in the SPN compared to the CSPN (219 vs. 150 spikes plant^{-1}). Plant height, median growth stage and stem angle were similar between the two environments, with CSPN plants growing slightly taller (Table 5.3). Seed yield in the CSPN environment was almost half of that observed in the SPN environment (6.53 vs 10.47 g plant^{-1} respectively).

It is important to note that the maximum and minimum range for crown rust severity are less than a 10% of what was found in either sward environment. It is probable that the microclimate around each plant in each of these growing environments had a heavy impact on the ability of crown rust to successfully infect the host. For instance, as plant density increased from 3 plants m^{-2} (SPN) to 5000 plants m^{-2} in the (TGSward) mean severity increased more than 100-fold from 0.01% to 1.6%. Increases in diseases such as crown rust due to plant density is axiomatic in epidemiology and is caused by increased inoculum load and moisture within the plant canopy (Burdon and Chilvers, 1982). However, seldom has this very applicable principle of epidemiology been applied to the experimental design of turfgrass breeding trials and nurseries.

Nurseries were randomized which should have limited spatial effects of winter damage. There was likely an environmental advantage to having plant closer together. Possibly there was more consistent snow cover associated with having plants closer together. The differences in seed yield between the two nursery designs was variable, but not unexpected. Winter kill is the likely reason for the slightly higher seed yields in the CSPN environment at Roseau. At St. Paul where there was relatively lower winter kill in the SPN compared to Roseau, and these plants produced far more seed and had a much wider variance (Table 5.3). The largest entry mean for crown width was 14.7 cm in the CSPN and 20.9 in the SPN. This is similar to what Waldron et al. (2008) found and also signifies that on average plant crowns were not in direct contact in the establishment year, although the canopies certainly overlapped. This was also likely why biomass was more difficult to predict using green pixel accumulation as leaf area index would be increased in the competitive environment.

Spaced plants vs. TGSward

The ability of spaced plants to predict turfgrass quality was dependent on the location in the establishment year. At Roseau, neither SPN nor CSPN quality scores were correlated with TGSward scores on an entry mean basis. However, at St. Paul TGSward scores were correlated with both CPSN (r_p = 0.75, $P < 0.001$) and SPN (r_p = 0.49, $P < 0.05$) environments (Table 5.2). Although the CSPN environment had a larger correlation, only slightly more correct selections were made compared to the SPN environment. Both spaced plant environments were able to significantly predict genetic color in the TGSward with the CSPN having a small advantage at Roseau. Correlation coefficients were over 0.80 between CSPN and TGSward for crown rust severity at both locations, however SPN only had a significant relationship at St. Paul (Table 5.2). Turf density in the establishing year was not consistently predictable by any spaced plant design using either crown width or green pixel accumulation (data not shown). Winter survival in the TGSward could be predicted using the CSPN or SPN environments at St. Paul only, SPN had a slight advantage in making correct selections. Turfgrass quality in the production year was highly correlated with winter survival and so was not included in the analysis. Vertical growth was consistently predicted by the CSPN environment at both Roseau (r_p = 0.42, P = 0.04) and St. Paul (r_p = 0.77, $P < 0.001$). Stemminess in TGSward was not predicted by spaced plant fertile tiller number at either location or spaced plant design. For turfgrass related traits, the CSPN environment achieved more significant and generally higher rank correlations estimates. However, out of 112 total selections, the CSPN environment made only 46 compared to the 41 correct SPN selections.

Spaced plants vs. SPSward

Biomass accumulation in the SPSward environment was consistently predicted only by the CSPN environment (Table 5.3). Crown rust severity in both spaced plant nurseries were significantly correlated with sward severity. It is interesting to note that CSPN had a much wider range of severities compared to SPN (0-0.54 vs. 0-0.15% respectively). For this trait the competitive environment also had more correct selections and higher rank correlation estimates. Winter survival in the SPSward was not consistently

predicted by either spaced plant environments. Growth stage, plant height and stem angle were both consistently predicted by CSPN, although the correlations were not always superior to SPN (Table 5.3). Fertile tiller number in the sward was never predicted by spaced plants ($P > 0.10$). Sward seed yield was significantly correlated with plant yield in the CSPN environment at both Roseau ($r_p = 0.72$, $P < 0.001$) and St. Paul ($r_p = 0.65$, $P = 0.01$) (Table 5.3). Similar to biomass, sward yields were not correlated with SPN plant yield. Yield in the noncompetitive environment was both larger and had more variance than in the competitive environment for biomass and seed yield (Table 5.3). This wider variance has been a common observation by many other authors for both biomass and seed yield which seems to lead to inflated response to selection (Stratton and Ohm, 1989; Samuel et al., 1970; Bugge, 1987; Bruckner et al., 1991). For seed production related traits, the CSPN environment almost always achieved more and larger rank correlations estimates. Substantially more correct selections were made using the CSPN with 57 correct selections out of 112 total compared to only 44 in the SPN design. Waldron et al. (2008) found similar numbers of correct selections in forage-type tall fescue with non-competitive designs.

Utility of the competitive design

Observing related populations for turfgrass quality traits was far easier when individuals were planted in close proximity to each other. For instance, several members of a related population could be assessed for merit against each other. One downside to this nested structure is that this may lead to unintentional bias and make statistical analysis more difficult. Also, harvesting seed was made more convenient because workers could sit and harvest 10 plants without having to repeatedly bend over. Moreover, image collection is far easier when plants are spaced closely together. Each SPN nursery on average, took 57 minutes to image, whereas the CSPN nurseries only took 23 minutes--a substantial savings in time. The closer spacing also means that more plants can fit in a given field space.

Some traits were difficult to observe in the competitive design, notably plant height and lodging. As the spaced plant's fertile tillers intertwined, they were difficult to observe on a plant by plant basis. There is also a slight disadvantage to using competitive designs

if winter kill or any other highly localized stress is imposed, as plants of the same population can all be negatively impacted. This may lead to bias in the results; however, in this study the CSPN design resulted in tiller survival that was more similar to those observed in the sward environments. An augmented nursery design and spatial adjustments could be implemented within the context of the CSPN design that would allow for more control of highly variable abiotic stresses such as winter kill (Casler and Brummer, 2008).

Conclusions

Although not superior in all respects, the increase in plant density from 3 to 23 plants m^{-2} increased the predictive ability of many traits. This was especially true for seed production related traits. Image analysis was deployed successfully for some traits including crown rust severity, winter survival, and stemminess in the spaced plant and sward environments. Although these results seem promising future research should use a half-sib mating designs to calculate genetic correlations and selection efficiency to further explore the utility in increasing competition in nursery environments. Also, additional years of data will be needed to understand if selections at St. Paul can predict phenotypes at Roseau for seed yield capacity and winter survival.

Tables

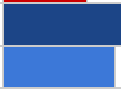
Table 5.1 – Systematic literature review of sources reporting the predictive ability of spaced plants for sward environments. Sward environments defined as managed grass plots or fields used for seed production, forage, or turfgrass.

Grass Species	Env. 1	Env. 2	Predictive ability	Source
Tall fescue	Spaced plants	SPS	Poor	Waldron et al. (2008)
Perennial ryegrass	Spaced plants	SPS	Poor	Elgersma (1990)
Perennial ryegrass	Space plants	SPS	Poor	Elgersma (1994)
Perennial ryegrass	Space plants	BFS	Poor	Lazenby and Rogers (1964)
Perennial ryegrass	Space plants	BFS	Poor	Hayward and Vivero (1984)
Perennial ryegrass	Imitative drill	BFS	Poor	Hayward and Vivero (1984)
Perennial ryegrass	Space plants	BFS	Fair	Foster (1971a); Foster (1971b); Foster (1973)
Perennial ryegrass	Micro sward	BFS	Poor	Foster (1971a); Foster (1971b); Foster (1973)
Perennial ryegrass	Space plants	BFS	Fair	Lazenby (1957)
Perennial ryegrass	Drilled rows	BFS	Fair	Lazenby (1957)
Switchgrass	Space plant	BFS	Fair	Bhandari et al. (2013)
Cereal rye	Space plant	BFS	Fair	Bruckner et al. (1991)
Bahiagrass	Space plant	Small sward	Good	Burton (1982)
Cereal rye	Space plant	Space plants	Poor	Kyriakou and Fasoulas (1985)
Cereal rye	Space plant	Space plants	Poor	Kyriakou and Fasoulas (1985)
Switchgrass	Space plant	Simulated Sward	Poor	Sykes et al. (2016)
Perennial ryegrass	Space plant	BFS	Good	Sedcole and Clements (1973)
Perennial ryegrass	Space plant	SPS	Poor	Bugge (1987)
Orchardgrass	Space plant	SPS	Poor	Stratton and Ohm (1989)

Table 5.2 – Predictive ability of spaced plants for turfgrass related traits. Entry minimum, maximum, mean, and standard deviation is provided for each location and spaced plant environment. Spearman rank correlation estimates are provided with the number of correct selections. Correct selections for Roseau (red) and St. Paul (blue) were determined by ordering entries in the sward environment and then selecting the top and bottom 20%.

			min	mean	max	sd	r_s	Correct
	Environment							1 2 3 4 5 6 7 8
Turf quality	Roseau	CSPN	3.3	6.2	7.6	0.8	NS	
		SPN	3.6	6.1	7.0	0.6	NS	
	St. Paul	CSPN	3.2	6.0	7.1	0.7	0.75	
		SPN	3.1	5.8	6.6	0.7	0.49 *	
Genetic color	Roseau	CSPN	4.0	6.5	7.6	0.6	0.63	
		SPN	4.9	6.9	7.8	0.6	0.49 *	
	St. Paul	CSPN	5.1	6.6	7.3	0.4	0.74	
		SPN	4.8	6.8	7.3	0.5	0.78	
Crown rust	Roseau	CSPN	0.0000	0.0001	0.0018	0.0002	0.82	
		SPN	0.0000	0.0001	0.0008	0.0002	NS	
	St. Paul	CSPN	0.0000	0.0010	0.0072	0.0013	0.81	
		SPN	0.0000	0.0003	0.0022	0.0004	0.74	
Density / Crown width	Roseau	CSPN	6.1	7.7	9.4	0.6	NS	
		SPN	6.8	7.9	9.4	0.6	NS	
	St. Paul	CSPN	6.3	7.7	9.5	0.6	0.67	
		SPN	7.1	8.7	10.7	0.7	NS	
Winter survival	Roseau	CSPN	0.0000	0.2234	0.5791	0.1707	NS	
		SPN	0.0053	0.0493	0.1290	0.0278	NS	
	St. Paul	CSPN	0.0035	0.4578	0.7720	0.2302	0.44 *	
		SPN	0.1762	0.3641	0.6386	0.1397	0.61	
Vertical growth	Roseau	CSPN	7.6	29.7	46.7	6.9	0.43 *	
		SPN	18.6	28.0	44.9	4.0	NS	
	St. Paul	CSPN	20.5	39.5	63.8	6.6	0.77	
		SPN	28.3	35.2	57.1	5.0	0.57	
Stemminess / Fertile tilers	Roseau	CSPN	0	108	249	56	NS	
		SPN	5	76	173	42	NS	
	St. Paul	CSPN	26	150	271	49	NS	
		SPN	115	219	386	56	NS	

Table 5.3 – Predictive ability of spaced plants for seed production related traits. Entry minimum, maximum, mean, and standard deviation is provided for each location and spaced plant environment. Spearman rank correlation estimates are provided with the number of correct selections. Correct selections for Roseau (red) and St. Paul (blue) were determined by ordering entries in the sward environment and then selecting the top and bottom 20%.

	Environment		min	mean	max	sd	r_s	Correct selections
								1 2 3 4 5 6 7 8
Biomass	Roseau	CSPN	11.8	30.9	57.9	8.9	0.46 *	
		SPN	32.3	43.5	58.7	6.7	NS	
	St. Paul	CSPN	26.4	50.1	95.8	13.4	0.77 ***	
		SPN	45.7	85.4	163.8	26.7	NS	
Crown rust	Roseau	CSPN	0.0000	0.0001	0.0018	0.0002	0.86 ***	
		SPN	0.0000	0.0001	0.0008	0.0002	0.54 **	
	St. Paul	CSPN	0.0000	0.0010	0.0072	0.0013	0.93 ***	
		SPN	0.0000	0.0003	0.0022	0.0004	0.85 ***	
Winter survival	Roseau	CSPN	0.00	0.22	0.58	0.17	NS	
		SPN	0.01	0.05	0.13	0.03	NS	
	St. Paul	CSPN	0.00	0.46	0.77	0.23	0.51 *	
		SPN	0.18	0.36	0.64	0.14	NS	
Stem angle	Roseau	CSPN	4.0	5.9	7.0	0.7	0.52 **	
		SPN	3.5	5.0	6.0	0.6	NS	
	St. Paul	CSPN	1.0	5.2	7.0	1.6	0.49 **	
		SPN	4.2	5.1	6.1	0.4	0.63 **	
Height	Roseau	CSPN	7.6	29.7	46.7	6.9	0.43 *	
		SPN	18.6	28.0	44.9	4.0	NS	
	St. Paul	CSPN	20.5	39.5	63.8	6.6	0.55 **	
		SPN	28.3	35.2	57.1	5.0	0.60 **	
Seed yield	Roseau	CSPN	0.00	2.81	10.29	2.05	0.72 ***	
		SPN	0.17	2.72	13.86	2.24	NS	
	St. Paul	CSPN	0.41	6.53	12.08	2.55	0.65 **	
		SPN	4.65	10.47	22.62	4.00	NS	
Growth stage	Roseau	CSPN	39	45	59	-	0.52 *	
		SPN	31	45	50	-	NS	
	St. Paul	CSPN	45	50	61	-	0.57 **	
		SPN	45	50	61	-	.45 *	

Figures

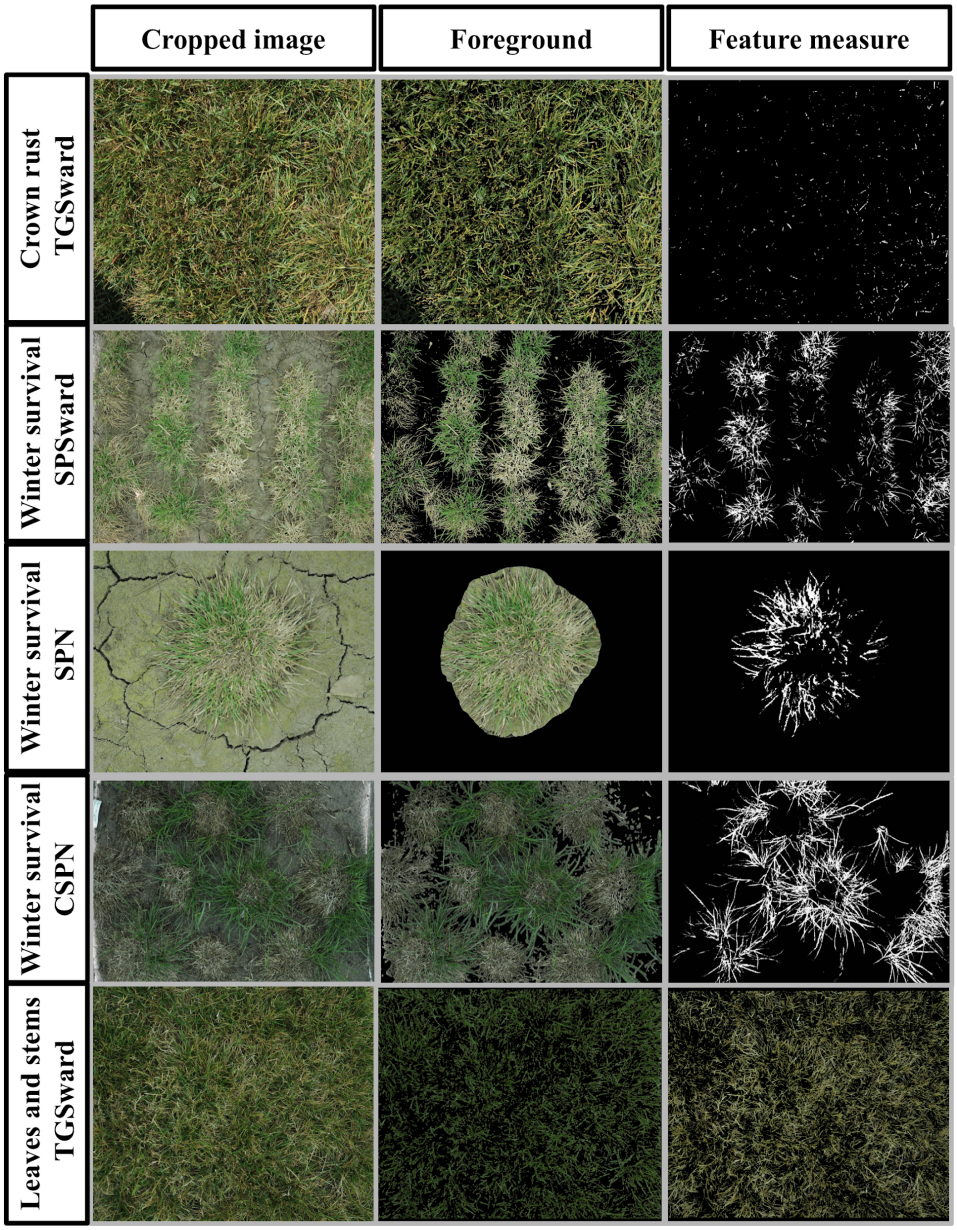


Figure 5.1 - Diagram displaying features from the four selection environments that were quantified using image analysis. The first column shows the cropped image, which is the first step in the image analysis pipeline. The second column displays foreground prediction and the result of the original RGB values from the cropped image being overlaid. This step allows for quality control of each feature measured. The third column shows the predicted feature from which the quantitative trait value is derived from. Each row displays a different trait in including crown rust, winter survival in SPN, CSPN, and SPSward as well as healthy and stem tissue in TGSward.

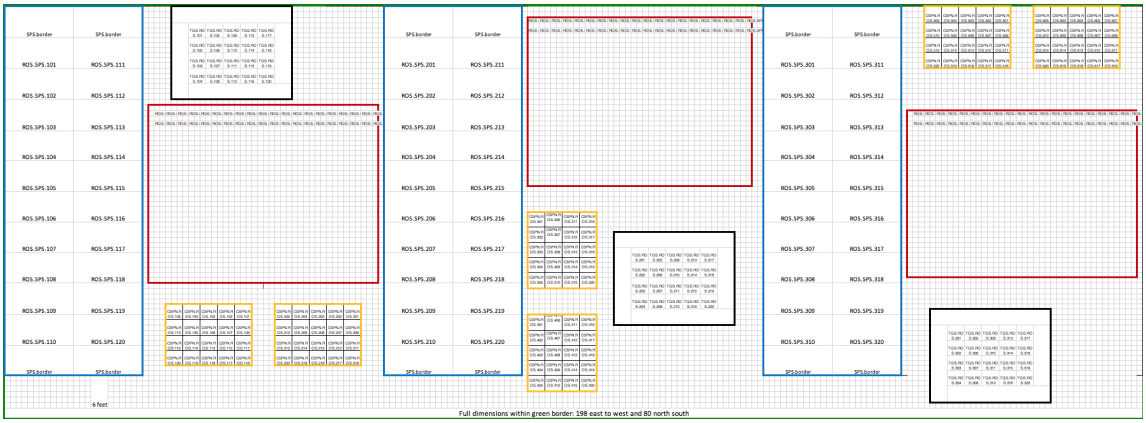


Figure 5.2 - Field plan displaying the configuration of the four growing environments: spaced plant nursery (SPN) is outlined in red, competitive spaced plant nursery (CSPN) outlined in orange, seed production sward (SPSward) outlined in blue, and turfgrass (TGS) plots are outlined in black. Each block includes a randomized arrangement of all three growing environments.

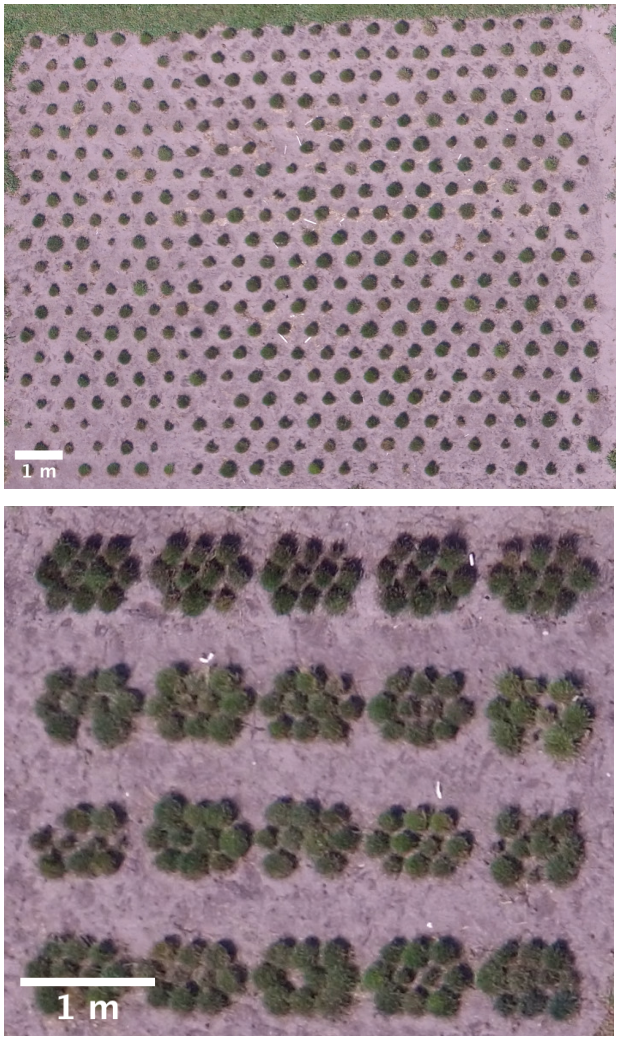


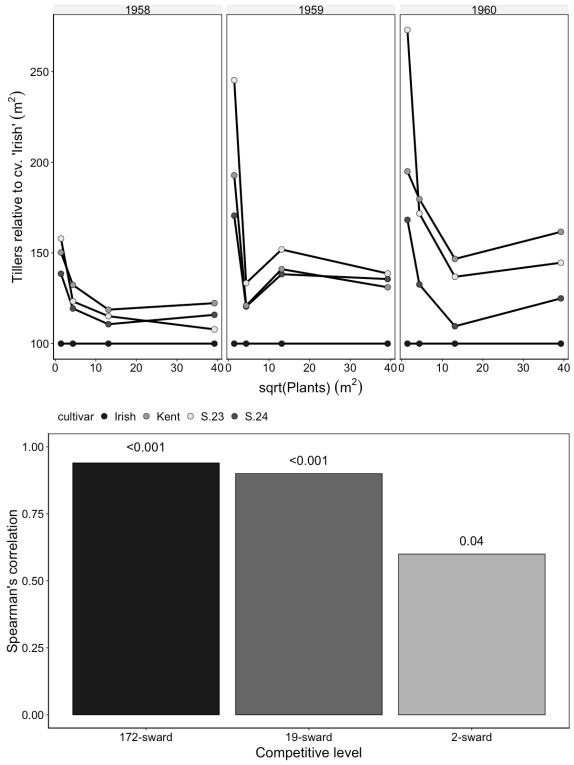
Figure 5.3 - Images of a single block of both CSPN and SPN growing environments. Standard spaced plant block (top) consisted of 400 plants 0.61 m apart arranged in a 20x20 Latin square. Competitive spaced plant (bottom) included 20 competitive units each made up of 10 plants 0.2 m apart from the same entry.

Supplemental Tables

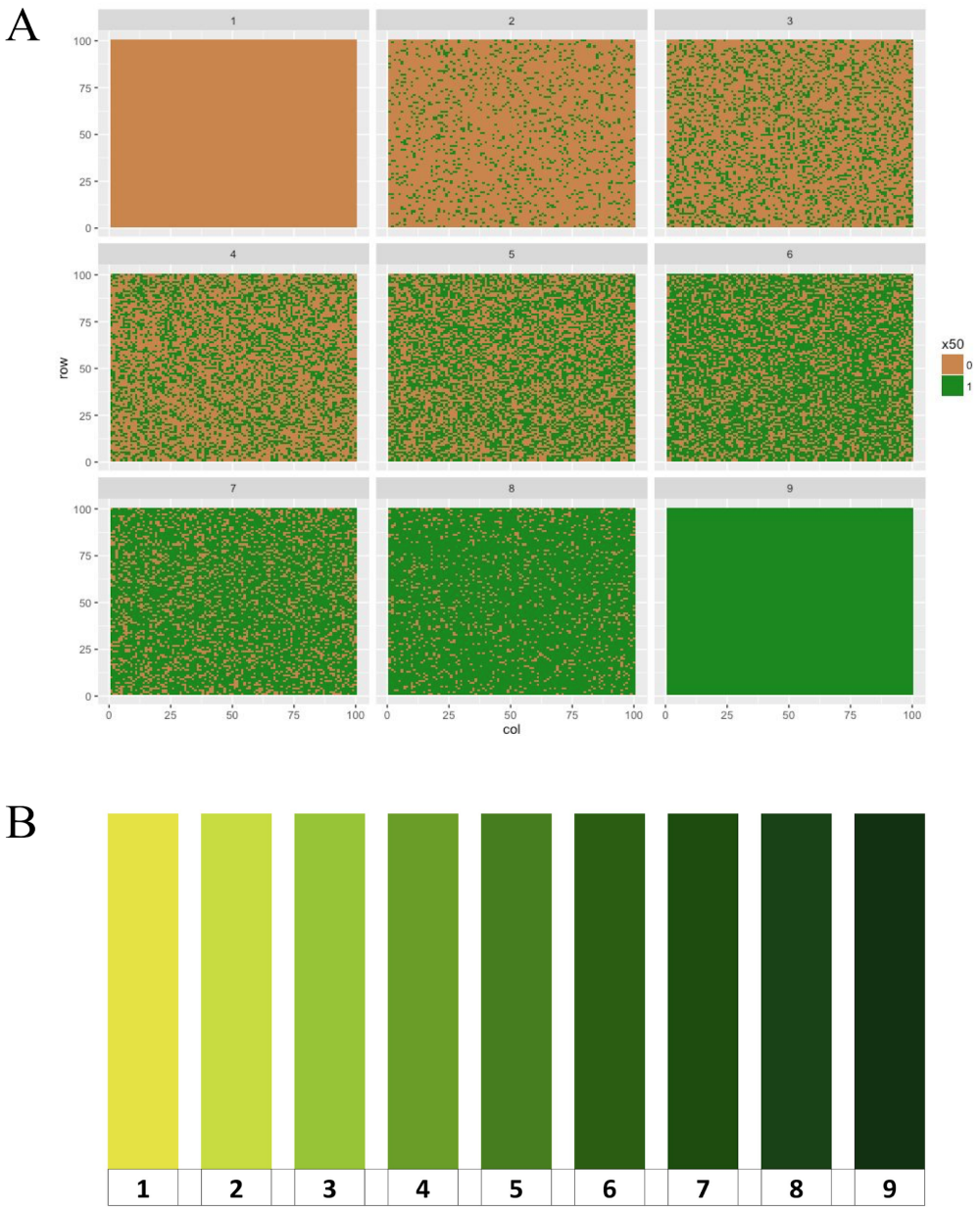
Supplemental Table 1 - Experimental entries included in both the turfgrass and seed production trials. Check entries were commercial cultivars and include information on turfgrass quality, seed yield, and winter hardiness. Experimental entries were interrelated half-sib populations contributed by DLF International Seeds breeding program with parental sources originating from commercial cultivars and advanced breeding populations in Minnesota and Connecticut.

Entry designation	Cultivar/selection name	Germplasm type	Seed source	Parental source	Turfgrass quality rating †	Relative seed yield 2 ‡	Winter hardiness §
PRSM.17.e01	Quebec	Check	DLF	NA	3.1	NA	-2.7
PRSM.17.e02	NK 200	Check	University of MN	NA	NA	Poor	0
PRSM.17.e03	Metolius	Check	Peak Plant Genetics	NA	6.1	NA	NA
PRSM.17.e04	Fiesta_4	Check	DLF	NA	4.7	Moderate	-2.4
PRSM.17.e05	Arctic_Green	Check	University of MN	NA	4.3	Excellent	-0.7
PRSM.17.e06	NGSPRWH-2-12	Experimental	OR	NA	NA	NA	NA
PRSM.17.e07	PR-24-16-06	Experimental	Connecticut	PSG.1037-12K	NA	NA	NA
PRSM.17.e08	PR-24-16-08	Experimental	Connecticut	Zoom	NA	NA	NA
PRSM.17.e09	PR-24-16-09	Experimental	Minnesota	PR.580	NA	NA	NA
PRSM.17.e10	PR-24-16-12	Experimental	Connecticut	Bandalore	NA	NA	NA
PRSM.17.e11	PR-24-16-14	Experimental	Connecticut	PR.562	NA	NA	NA
PRSM.17.e12	PR-24-16-17	Experimental	Minnesota	Karma	NA	NA	NA
PRSM.17.e13	PR-24-16-21	Experimental	Minnesota	PR.524C	NA	NA	NA
PRSM.17.e14	PR-24-16-22	Experimental	Connecticut	Diligent	NA	NA	NA
PRSM.17.e15	PR-24-16-23	Experimental	Minnesota	Monsieur	NA	NA	NA
PRSM.17.e16	PR-24-16-25	Experimental	Minnesota	Aspire	NA	NA	NA
PRSM.17.e17	PR-24-16-26	Experimental	Connecticut	PSG.HLT	NA	NA	NA
PRSM.17.e18	PR-24-16-28	Experimental	Connecticut	PR.583	NA	NA	NA
PRSM.17.e19	PRWHNGS-1-14	Experimental	OR	NA	NA	NA	NA
PRSM.17.e20	WH-PR-11-03	Experimental	OR	NA	NA	NA	NA
† Turfgrass quality determined by national turfgrass trial data.							
‡ Seed production capacity determined by MTSC reports relative to NK200 (2013-2017).							
§ Winter survival as reported by Minnesota Turf Seed Council in relation to NK200 in 2016 (1 unit = 12.5 % change in tiller survival).							

Supplemental Figures



Supplemental Figure 5.1 - Data adapted from Lazenby and Rogers (1964) Table 7. Top panel shows relative tiller number for four cultivars compared to the cultivar Irish across three years. Plant density was transformed for ease of observation, exact plant densities were 2, 19, 172, 1540 plants m^{-2}. Bottom panel shows the spearman's rank correlations of each space plant environment compared to the sward (1540 plants m^{-2}) with *P* values denoted above columns.



Supplemental Figure 5.2 - Diagrammatic rating scales used for measuring lateral growth and genetic turfgrass color. A) Each panel represents a 12.5% increase in green cover for each of the nine density categories. B) Custom color palette for rating genetic color. Scale values (HSB): 1 = 60-100-90, 2 = 67-100-86, 3 = 74-100-86, 4 = 81-100-62, 5 = 88=100=50, 6 = 96=100-37, 7 = 102-100-31, 8 = 111-90-26, 9 = 120-90-20.

References

Aamlid, T.S., O.M. Heide, and B. Boelt. 2000. Primary and secondary induction requirements for flowering of contrasting European varieties of *Lolium perenne*. Annals of Botany. 86(6): 1087–1095.

Abel, S., R. Gislum, and B. Boelt. 2017. Path and correlation analysis of perennial ryegrass (*Lolium perenne* L.) seed yield components. Journal of Agronomy and Crop Science. doi/10.1111/jac.12202/full.

Ahlgren, H.L., D.C. Smith, and E.L. Nielsen. 1945. Behavior of various selections of Kentucky bluegrass, *Poa pratensis* L., when grown as spaced plants and in mass seedings. Journal of the American Society of Agronomy.

Alonso, M.P., N.E. Mirabella, J.S. Panelo, M.G. Cendoya, and A.C. Pontaroli. 2018. Selection for high spike fertility index increases genetic progress in grain yield and stability in bread wheat. Euphytica. 214(7): 112.

Armstead, I.P., L.B. Turner, A.H. Marshall, M.O. Humphreys, I.P. King, et al. 2008. Identifying genetic components controlling fertility in the outcrossing grass species perennial ryegrass (*Lolium perenne*) by quantitative trait loci analysis and comparative genetics. New Phytologist. 178(3): 559–571.

Bacon, C.W., and J.F. White. 2016. Functions, mechanisms and regulation of endophytic and epiphytic microbial communities of plants. Symbiosis. 68(1–3): 87–98.

Bade, C.I.A., and M.A. Carmona. 2011. Comparison of methods to assess severity of common rust caused by *Puccinia sorghi* in maize. Tropical Plant Pathology. 36(4): 264–266.

Bai, G., S. Jenkins, W. Yuan, G.L. Graef, and Y. Ge. 2018. Field-based scoring of soybean iron deficiency chlorosis using RGB imaging and statistical learning. Frontiers in Plant Science 9.

Bastias, D.A., M.A. Martínez-Ghersa, C.L. Ballaré, and P.E. Gundel. 2017. *Epichloë* Fungal endophytes and plant defenses: Not just alkaloids. Trends in Plant Science. 22(11), 939-948.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using {lme4}. Journal of Statistical Software 67(1): 1–48.

Becker, M., Y. Becker, K. Green, and B. Scott. 2016. The endophytic symbiont *Epichloë festucae* establishes an epiphyllous net on the surface of *Lolium perenne* leaves by development of an expressorium, an appressorium-like leaf exit structure. New Phytologist. doi/10.1111/nph.13931/full.

Bernardo, R. 2016. Bandwagons I, too, have known. Theoretical and Applied Genetics. 129(12): 2323–2332.

Bolaric, S., S. Barth, A.E. Melchinger, and U.K. Posselt. 2005. Genetic diversity in European perennial ryegrass cultivars investigated with RAPD markers. Plant Breeding. 124(2): 161–166.

Bonos, S.A., and D.R. Huff. 2013. Cool-season grasses: Biology and breeding. Turfgrass: Biology, Use, and Management. 591–660.

Bonos, S.A., M.M. Wilson, W.A. Meyer, and C. Reed Funk. 2005. Suppression of red thread in fine fescues through endophyte-mediated resistance. Applied Turfgrass Science. 2(1): 0–0.

Bruckner, P.L., P.L. Raymer, and G.W. Burton. 1991. Recurrent phenotypic selection for forage yield in rye. Euphytica. 54(1): 11–17.

Bugge, G. 1987. Selection for seed yield in *Lolium perenne* L. Plant Breeding. 98(2): 149–155.

Burbidge, A., P.D. Hebblethwaite, and J.D. Ivins. 1978. Lodging studies in *Lolium perenne* grown for seed: 2. Floret site utilization. The Journal of Agricultural Science. 90(2): 269–274.

Burdon, J.J., and G.A. Chilvers. 1982. Host density as a factor in plant disease ecology. Annual Review of Phytopathologyi 20(1): 143–166.

Burton, G.W. 1982. Improved recurrent restricted phenotypic selection increases bahiagrass forage yields. Crop Science 22(5): 1058–1061.

Bylin, A.G., S.D. Card, D.E. Hume, and C.M. Lloyd-West. 2016. Endophyte storage and seed germination of *Epichloë*-infected meadow fescue. Seed Science and Technology. 44(1): 138–155.

Byrne, S., E. Guiney, S. Barth, I. Donnison, L.A. Mur, et al. 2009. Identification of coincident QTL for days to heading, spike length and spikelets per spike in *Lolium perenne* L. Euphytica. 166(1): 61–70.

Capella-Gutiérrez, S., J.M. Silla-Martínez, and T. Gabaldón. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25(15): 1972–1973.

Casler, M.D. 2006. Perennial grasses for turf, sport and amenity uses: evolution of form, function and fitness for human benefit. The Journal of Agricultural Science. 144(03): 189–203.

Casler, M.D. 2010. Changes in mean and genetic variance during two cycles of within-family selection in switchgrass. BioEnergy Research. 3(1): 47–54.

Casler, M.D., R.E. Barker, J.H. Cherney, and Y.A. Papadopolous. 2004. Stability of nonflowering orchardgrass. Crop Science. 44(5): 1601–1607.

Casler, M.D., and E.C. Brummer. 2008. Theoretical expected genetic gains for among-and-within-family selection methods in perennial forage crops. Crop Science. 48(3): 890–902.

Casler, M.D., J.F. Pedersen, G.C. Eizenga, and S.D. Stratton. 1996. Germplasm and cultivar development. Cool-Season Forage Grasses. 413–469.

Casler, M.D., and E. van Santen. 2008. Fungal endophyte removal does not reduce cold tolerance of tall fescue. Crop Science. 48(5): 2033–2039.

Casler, M.D., and E. Van Santen. 2010. Breeding objectives in forages. Fodder Crops and Amenity Grasses. Springer. p. 115–136

Chastain, T.G., and D.F. Grabe. 1989a. Spring establishment of orchardgrass seed crops with cereal companion crops. Crop Science 29(2): 466–471.

Chastain, T.G., and D.F. Grabe. 1989b. Spring establishment of turf-type tall fescue seed crops with cereal companion crops. Agronomy Journal. 81(3): 488–493.

Chastain, T.G., and W.C. Young III. 1998. Vegetative plant development and seed production in cool-season perennial grasses. Seed Science Research. 8(2): 295–302.

Christensen, M.J., and G.C.M. Latch. 1991. Variation among isolates of Acremonium endophytes (*A. coenophialum* and possibly *A. typhinum*) from tall fescue (*Festuca arundinacea*). Mycological Research. 95(9): 1123–1126.

Cicchetti, D.V. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 6(4): 284.

Clarke, R.G., and D.R. Eagling. 1994. Effects of pathogens on perennial pasture grasses. New Zealand Journal of Agricultural Research. 37(3): 319–327.

Clarke, B.B., J.F. White Jr, R.H. Hurley, M.S. Torres, S. Sun, et al. 2006. Endophyte-mediated suppression of dollar spot disease in fine fescues. Plant Disease. 90(8): 994–998.

Clay, K. 1989. Clavicipitaceous endophytes of grasses: their potential as biocontrol agents. Mycological Research. 92(1): 1–12.

Clay, K., and C. Schardl. 2002. Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. The American Naturalist. 160(S4): S99–S127.

COLVILL, K.E., and C. Marshall. 1984. Tiller dynamics and assimilate partitioning in *Lolium perenne* with particular reference to flowering. Annals of Applied Biology. 104(3): 543–557.

Cooper, J.K., A.M.H. Ibrahim, J. Rudd, S. Malla, D.B. Hays, et al. 2012. Increasing hard winter wheat yield potential via synthetic wheat: I. Path-coefficient analysis of yield and its components. Crop Science. 52(5): 2014–2022.

Dewey, D.R., and K.H. Lu. 1959. A correlation and path-coefficient analysis of components of crested wheatgrass seed production. Agronomy Journal. 51(9): 515–518.

Diaz-Lago, J.E., D.D. Stuthman, and T.E. Abadie. 2002. Recurrent selection for partial resistance to crown rust in oat. Crop Science. 42(5): 1475–1482.

Díaz-Lago, J.E., D.D. Stuthman, and K.J. Leonard. 2003. Evaluation of components of partial resistance to oat crown rust using digital image analysis. Plant Disease. 87(6): 667–674.

Dofing, S.M., and C.W. Knight. 1992. Alternative model for path analysis of small-grain yield. Crop Science. 32(2): 487–489.

Donald, C.M. 1954. Competition among pasture plants. II. The influence of density on flowering and seed production in annual pasture plants. Australian Journal of Agricultural Research. 5(4): 585–597.

Donald, C.M. t. 1968. The breeding of crop ideotypes. Euphytica. 17(3): 385–403.

Dupont, P.-Y., C.J. Eaton, J.J. Wargent, S. Fechtner, P. Solomon, et al. 2015. Fungal endophyte infection of ryegrass reprograms host metabolism and alters development. New Phytologist. doi/10.1111/nph.13614/full.

Ehlke, N.J., and D.J. Vellekson. 1998. Progress Report on Seed Production Research. http://www.mnturfseed.org/html/progress_reports.html.

Elgersma, A. 1985. Floret site utilization in grasses: definitions, breeding perspectives and methodology. Journal of Applied Seed Production. 3: 50–54.

Elgersma, A. 1990a. Spaced-plant traits related to seed yield in plots of perennial ryegrass (*Lolium perenne* L.). Euphytica. 51(2): 151–161.

Elgersma, A. 1990b. Seed yield related to crop development and to yield components in nine cultivars of perennial ryegrass (*Lolium perenne* L.). Euphytica. 49(2): 141–154.

Elgersma, A. 1990c. Genetic, cytological and physiological aspects of seed yield in perennial ryegrass (*Lolium perenne* L.). Doctoral Dissertation. Retrieved from Wageningen University and Research Database.

Elgersma, A. 1990d. Heritability estimates of spaced-plant traits in three perennial ryegrass (*Lolium perenne* L.) cultivars. Euphytica. 51(2): 163–171.

Elgersma, A., and R. Śnieżko. 1988. Cytology of seed development related to floret position in perennial ryegrass (*Lolium perenne* L.). Euphytica. 39: 59–68.

Elgersma, A., G.D. Winkelhorst, and A.P.M. Den Nijs. 1994. The relationship between progeny seed yield in drilled plots and maternal spaced-plant traits in perennial ryegrass (*Lolium perenne* L.). Plant Breeding. 112(3): 209–214.

Elling, L. 1979. Progress report of seed production research. http://www.mnturfseed.org/html/progress_reports.html.

Elling, L. 1980. Progress report of seed production research. http://www.mnturfseed.org/html/progress_reports.html.

Florea, S., C.L. Schardl, and W. Hollin. 2015. Detection and Isolation of *Epichloë* Species, Fungal Endophytes of Grasses. Current Protocols in Microbiology. 19A–1.

Ford, V.L., and T.L. Kirkpatrick. 1989. Effects of *Acremonium coenophialu*m in tall fescue on host disease and insect resistance and allelopathy to *Pinus taeda* seedlings. Proceedings of the Arkansas Fescue Toxicosis Conference.

Foster, C.A. 1971. Interpopulational and intervarietal hybridization in *Lolium perenne* breeding: heterosis under non-competitive conditions. The Journal of Agricultural Science. 76(01): 107–130.

Foster, C.A. 1973. Interpopulational and intervarietal F1 hybrids in *Lolium perenne*: performance in field sward conditions. The Journal of Agricultural Science. 80(03): 463–477.

Fradgley, N.S., H.E. Creissen, H. Pearce, S.A. Howlett, B.D. Pearce, et al. 2017. Weed suppression and tolerance in winter oats. Weed Technology. 31(5): 740–751.

Gamer, M., J. Lemon, and I.F.P. Singh. 2015. IRR: various coefficients of interrater reliability and agreement. 2012. R package version 0.84.

Gehan, M.A., N. Fahlgren, A. Abbasi, J.C. Berry, S.T. Callen, et al. 2017. PlantCV v2: Image analysis software for high-throughput plant phenotyping. PeerJ 5. e4088.

Ghorbanpour, M., M. Omidvari, P. Abbaszadeh-Dahaji, R. Omidvar, and K. Kariman. 2017. Mechanisms underlying the protective effects of beneficial fungi against plant diseases. Biological Control. 117, pp.147-157.

Giauque, H., E.W. Connor, and C.V. Hawkes. 2018. Endophyte traits relevant to stress tolerance, resource use and habitat of origin predict effects on host plants. New Phytologist. 221(4), p.2239-2249.

Grace, J.B. 2006. Structural equation modeling and natural systems. Cambridge University Press.

GREULICH, F., 堀尾枝美子, 島貫忠幸, and 吉原照彦. 1999. Field results confirm natural plant protection by the endophytic fungus *Epichloe typhina* against the

pathogenic fungus *Cladosporium phlei* on timothy leaves. Japanese Journal of Phytopathology. 65(4), 454-459.

Gustavsson, A.M. 2011. A developmental scale for perennial forage grasses based on the decimal code framework. Grass and Forage Science. 66(1): 93–108.

Gwinn, K.D., and A.M. Gavin. 1992. Relationship between endophyte infestation level of tall fescue seed lots and *Rhizoctonia zeae* seedling disease. Plant Disease. 76(9): 911–914.

Hampton, J.G., and P.D. Hebblethwaite. 1983. The effects of the environment at anthesis on the seed yield and yield components of perennial ryegrass (*Lolium perenne* L.) cv. S 24. Journal of Applied Seed Production. 1: 21–22.

Hart, J.M., N.P. Anderson, A.G. Hulting, T.G. Chastain, M.E. Mellbye, et al. 2012. Postharvest residue management for grass seed production in western Oregon. Corvallis, Or. Extension Service, Oregon State University.

Hayward, M.D., and J.L. Vivero. 1984. Selection for yield in *Lolium perenne*. II. Performance of spaced plant selections under competitive conditions. Euphytica. 33(3): 787–800.

Hazard, L., M. Betin, and N. Molinari. 2006. Correlated response in plant height and heading date to selection in perennial ryegrass populations. Agronomy Journal. 98(6): 1384–1391.

Heineck, G.C., E. Watkins, and N.J. Ehlke. 2018a. The Fungal *Endophyte Epichloë festucae* var. *lolii Does* Not Improve the Freezing Tolerance of Perennial Ryegrass. Crop Science. 58.4 (2018): 1788-1800.

Heineck, G., E. Watkins, and N.J. Ehlke. 2018b. Exploring Alternative Management Options for Multiyear Perennial Ryegrass Seed Production in Northern Minnesota. Crop Science. 58(1): 426–434.

Heineck, G.C., I.G. McNish, J.M. Jungers, E. Gilbert, E. Watkins. 2019. Using R-based image analysis to quantify rust on perennial ryegrass. The Plant Phenome Journal. 2:180010. doi:10.2135/tppj2018.12.0010.

Hesse, U., W. Schöberlein, L. Wittenmayer, K. Förster, K. Warnstorff, et al. 2003. Effects of *Neotyphodium* endophytes on growth, reproduction and drought-stress

tolerance of three *Lolium perenne* L. genotypes. Grass and Forage Science 58(4): 407–415.

Hettiarachchige, I.K., P.N. Ekanayake, R.C. Mann, K.M. Guthridge, T.I. Sawbridge, et al. 2015. Phylogenomics of asexual *Epichloë* fungal endophytes forming associations with perennial ryegrass. BMC Evolutionary Biology. 15(1): 72.

Hiatt, E.E., N.S. Hill, J.H. Bouton, and C.W. Mims. 1997. Monoclonal antibodies for detection of *Neotyphodium coenophialum*. Crop Science. 37(4): 1265–1269.

Hill, E.C., K.A. Renner, C.L. Sprague, and J.E. Fry. 2017. Structural equation modeling of cover crop effects on soil nitrogen and dry bean. Agronomy Journal. 109(6): 2781–2788.

Hill, M.J., and B.R. Watkin. 1975. Seed production studies on perennial ryegrass, timothy and prairie grass. Grass and Forage Science. 30(1): 63–71.

Hu, L., and P.M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 6(1): 1–55.

Huerta-Cepas, J., F. Serra, and P. Bork. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Molecular Biology and Evolution. 33(6): 1635–1638.

Hulke, B.S., E. Watkins, D. Wyse, and N. Ehlke. 2007. Winterhardiness and turf quality of accessions of perennial ryegrass from public collections. Crop Science. 47(4): 1596–1602.

Humphreys, M., U. Feuerstein, M. Vandewalle, and J. Baert. 2010. Ryegrasses. Fodder Crops and Amenity Grasses. Springer. p. 211–260

Islam, M.A., A.K. Obour, J.M. Krall, J.T. Cecil, and J.J. Nachtman. 2013. Performance of turfgrass under supplemental irrigation and rain-fed conditions in the Central Great Plains of USA. Grassland Science. 59(2): 111–119.

Jessen, D.L., and I.T. Carlson. 1985. Response to selection for seed and forage traits in smooth bromegrass. Crop Science. 25(3): 502–505.

Johnson, L.J., A.C. de Bonth, L.R. Briggs, J.R. Caradus, S.C. Finch, et al. 2013. The exploitation of *Epichloae* endophytes for agricultural benefit. Fungal Diversity. 60(1): 171–188.

Johnson, R.C., W.J. Johnston, and C.T. Golob. 2003. Residue management, seed production, crop development, and turf quality in diverse Kentucky bluegrass germplasm. Crop Science. 43(3): 1091–1099.

Johnson-Cicalese, J., M.E. Secks, C.K. Lam, W.A. Meyer, J.A. 2000. Cross species inoculation of Chewings and strong creeping red fescues with fungal endophytes. Crop Science. 40(5): 1485–1489.

Johnston, W.J., K.L. Dodson, D.A. Silbernagel, G.K. Stahnke, R.C. 2011. Kentucky Bluegrass Seed Production without Field Burning while Maintaining. Turfgrass Quality. http://turf.wsu.edu/wp-content/uploads/2013/09/DOE-2009-2011-FINAL-REPORT-nbkbg.pdf.

Kane, K.H. 2011. Effects of endophyte infection on drought stress tolerance of *Lolium perenne* accessions from the Mediterranean region. Environmental and Experimental Botany. 71(3): 337–344.

Karcher, D.E., and M.D. Richardson. 2003. Quantifying turfgrass color using digital image analysis. Crop Science. 43(3): 943–951.

Karcher, D.E., and M.D. Richardson. 2005. Batch analysis of digital images to evaluate turfgrass characteristics. Crop Science. 45(4): 1536–1539.

Katoh, K., and D.M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution. 30(4): 772–780.

Koeritz, E.J., E. Watkins, and N.J. Ehlke. 2013. A split application approach to nitrogen and growth regulator management for perennial ryegrass seed production. Crop Science. 53(4): 1762–1777.

Koeritz, E.J., E. Watkins, and N.J. Ehlke. 2015. Seeding rate, row spacing, and nitrogen rate effects on perennial ryegrass seed production. Crop Science. 55(5): 2319–2333.

Kuhn, M. 2008. Building predictive models in R using the caret package. Journal of Statistical Software. 28(5): 1–26.

Kuldau, G., and C. Bacon. 2008. Clavicipitaceous endophytes: their ability to enhance resistance of grasses to multiple stresses. Biological Control. 46(1): 57–71.

Kurtzman, C.P., and C.J. Robnett. 1997. Identification of clinically important ascomycetous yeasts based on nucleotide divergence in the 5'end of the large-subunit (26S) ribosomal DNA gene. Journal of Clinical Microbiology. 35(5): 1216–1223.

Lamb, E., S. Shirtliffe, and W. May. 2011. Structural equation modeling in the plant sciences: An example using yield components in oat. Canadian Journal of Plant Science. 91(4): 603–619.

Lazenby, A., and H.H. Rogers. 1964. Selection criteria in grass breeding. II. Effect, on *Lolium perenne*, of differences in population density, variety and available moisture. The Journal of Agricultural Science. 62(02): 285–298.

Lenth, R. 2018. Emmeans: Estimated marginal means, aka least-squares means. R Package Version 1(2).

Leuchtmann, A., C.W. Bacon, C.L. Schardl, J.F. White, and M. Tadych. 2014. Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë.* Mycologia. 106(2): 202–215.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News. 2(3).

Lodge, G.M. 1993. The domestication of the native grasses *Danthonia richardsonii* Cashmore and *Danthonia linkii Kunth* for agricultural use. I. Selecting for inflorescence seed yield. Australian Journal of Agricultural Research. 44(1): 59–77.

Lumley, T., and M.T. Lumley. 2013. Package 'leaps.' regression subset selection. Thomas Lumley Based on Fortran Code by Alan Miller. http://CRAN. R-project. org/package= leaps.

Madden, L.V., G. Hughes, and F. Bosch. 2007. The study of plant disease epidemics. American Phytopathological Society. (APS Press).

Martino, D.L., P.E. Abbate, M.G. Cendoya, F. Gutheim, N.E. Mirabella, et al. 2015. Wheat spike fertility: inheritance and relationship with spike yield components in early generations. Plant Breeding. 134(3): 264–270.

McDonagh, J., M. O'Donovan, M. McEvoy, and T.J. Gilliland. 2016. Genetic gain in perennial ryegrass (*Lolium perenne*) varieties 1973 to 2013. Euphytica. 212(2): 187–199.

McDonald, M., and L.O. Copeland. 1997. Seed production: principles and practices. Chapman and Hall, New York, NY.

Mendiburu, F. 2016. agricolae: Statistical procedures for agricultural research.

Meyer, W.A., and C.R. Funk. 1989. Progress and benefits to humanity from breeding cool-season grasses for turf 1. Contributions from Breeding Forage and Turf Grasses. 31–48.

Mirabella, N.E., P.E. Abbate, I.A. Ramirez, and A.C. Pontaroli. 2016. Genetic variation for wheat spike fertility in cultivars and early breeding materials. The Journal of Agricultural Science. 154(1): 13–22.

Morris, K.M., and R.C. Shearman. 2006. NTEP turfgrass evaluation guidelines. National Turfgrass Evaluation Program.

Moy, M., F. Belanger, R. Duncan, A. Freehoff, C. Leary, et al. 2000. Identification of epiphyllous mycelial nets on leaves of grasses infected by clavicipitaceous endophytes. Symbiosis. 28(4): 291–302.

Naik, H.S., J. Zhang, A. Lofquist, T. Assefa, S. Sarkar, et al. 2017. A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. Plant Methods. 13(1): 23.

NASS. 2016. Oregon Agriculture Facts and Figures. https://www.nass.usda.gov/Statistics_by_State/Oregon/Publications/facts_and_figures/facts_and_figures.pdf.

NASS. 2018. Oregon agriculture facts and figures. https://www.nass.usda.gov/Statistics_by_State/Oregon/Publications/facts_and_figures/facts_and_figures.pdf.

Niones, J.T., and D. Takemoto. 2014. An isolate of *Epichloë festucae*, an endophytic fungus of temperate grasses, has growth inhibitory activity against selected grass pathogens. Journal of General Plant Pathology. 80(4): 337–347.

Onishi, K., Y. Sano, and H. Nakashima. 2003. Developmental fates of axillary buds as a major determinant for the pattern of life history in *Lolium*. Plant Production Science. 6(3): 179–184.

Pańka, D., M. Jeske, and M. Troczyński. 2011. Effect of *Neotyphodium uncinatum* endophyte on meadow fescue yielding, health status and ergovaline production in host-plants. Journal of Plant Protection Research. 51(4): 362–370.

Pańka, D., D. Piesik, M. Jeske, and A. Baturo-Cieśniewska. 2013. Production of phenolics and the emission of volatile organic compounds by perennial ryegrass (*Lolium perenne* L.)/*Neotyphodium lolii* association as a response to infection by *Fusarium poae*. Journal of Plant Physiology. 170(11): 1010–1019.

Panka, D., L. Podkówka, and R. Lamparski. 2004. Preliminary observations on the resistance of meadow fescue (*Festuca pratensis* Huds.) infected by *Neotyphodium uncinatum* to diseases and pests and native value. Proceedings of the 5th International Symposium on *Neotyphodium*/Grass Interactions.

Patrignani, A., and T.E. Ochsner. 2015. Canopeo: A Powerful New Tool for Measuring Fractional Green Canopy Cover. Agronomy Journal. 107(6): 2312–2320.

Pau, G., F. Fuchs, O. Sklyar, M. Boutros, and W. Huber. 2010. EBImage—an R package for image processing with applications to cellular phenotypes. Bioinformatics. 26(7): 979–981.

Pennekamp, F., and N. Schtickzelle. 2013. Implementing image analysis in laboratory-based experimental systems for ecology and evolution: a hands-on guide. Methods in Ecology and Evolution. 4(5): 483–492.

Perez, L.I., P.E. Gundel, H.J. Marrero, A.G. Arzac, and M. Omacini. 2017. Symbiosis with systemic fungal endophytes promotes host escape from vector-borne disease. Oecologia. 184(1): 237–245.

Pfender, W. 2009. A damage function for stem rust of perennial ryegrass seed crops. Phytopathology. 99(5): 498–505.

Philipp, N., H. Weichert, U. Bohra, W. Weschke, A.W. Schulthess, et al. 2018. Grain number and grain yield distribution along the spike remain stable despite breeding for high yield in winter wheat. PloS One. 13(10): e0205452.

Plummer, R.M., R.L. Hall, and T.A. Watt. 1990. The influence of crown rust (*Puccinia coronata*) on tiller production and survival of perennial ryegrass (*Lolium perenne*) plants in simulated swards. Grass and forage science 45(1): 9–16.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raeber, J.G., and R.R. Kalton. 1956. Variation and inheritance of fertility and its components in *Bromus inermis* Leyss. Agronomy Journal. 48(5): 212–216.

Rambaut, A. 2012. FigTree v1. 4.

Ravel, C., G. Charmet, and F. Balfourier. 1995. Influence of the fungal endophyte *Acremonium lolii* on agronomic traits of perennial ryegrass in France. Grass and Forage Science. 50(1): 75–80.

Reed, K.F.M., A. Leonforte, P.J. Cunningham, J.R. Walsh, D.I. Allen, et al. 2000. Incidence of ryegrass endophyte (*Neotyphodium lolii*) and diversity of associated alkaloid concentrations among naturalized populations of perennial ryegrass (*Lolium perenne* L.). Australian Journal of Agricultural Research. 51(5): 569–578.

Roelfs, A.P. 1992. Rust diseases of wheat: concepts and methods of disease management. Cimmyt.

Rolston, P., J. Trethewey, B. McCloy, and R. Chynoweth. 2007. Achieving forage ryegrass seed yields of 3000 kg/ha and limitations to higher yields. Proceedings of the 6th International Herbage Seed Conference. p. 100–106.

Rueden, C.T., J. Schindelin, M.C. Hiner, B.E. DeZonia, A.E. Walter, et al. 2017. ImageJ2: ImageJ for the next generation of scientific image data. BMC Bioinformatics. 18(1): 529.

Sampoux, J.P., P. Baudouin, B. Bayle, V. Béguier, P. Bourdon, et al. 2013. Breeding perennial ryegrass (*Lolium perenne* L.) for turf usage: an assessment of genetic improvements in cultivars released in Europe, 1974–2004. Grass and Forage Science. 68(1): 33–48.

Samuel, C.J.A., J. Hill, E.L. Breese, and A. Davies. 1970. Assessing and predicting environmental response in *Lolium perenne*. The Journal of Agricultural Science. 75(01): 1–9.

Sankar, M., K. Nieminen, L. Ragni, I. Xenarios, and C.S. Hardtke. 2014. Automated quantitative histology reveals vascular morphodynamics during *Arabidopsis* hypocotyl secondary growth. Elife. 3: e01567.

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, et al. 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods. 9(7): 676.

Schneider, C.A., W.S. Rasband, and K.W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 9(7): 671.

Schubiger, F.X., and B. Boller. 2016. Virulence of crown rust isolates (*Puccinia coronata* f. sp. *lolii*) on genotypes of Italian and perennial ryegrass (*Lolium multiflorum* and *L. perenne*). European Journal of Plant Pathology. 144(1): 141–154.

Schwanck, A.A., and E.M. Del Ponte. 2014. Accuracy and reliability of severity estimates using linear or logarithmic disease diagram sets in true colour or black and white: a study case for rice brown spot. Journal of Phytopathology. 162(10): 670–682.

Schwanck, A.A., and E.M. Del Ponte. 2016. Measuring lesion attributes and analyzing their spatial patterns at the leaf scale using digital image analysis. Plant Pathology. 65(9): 1498–1508.

Sedcole, J.R., and R.J. Clements. 1973. Studies on genotype-time-spacing interactions for herbage yield, using a modified diallel analysis. The Journal of Agricultural Science. 80(1): 97–104.

Sedgley, R.H. 1991. An appraisal of the Donald ideotype after 21 years. Field Crops Research. 26(2): 93–112.

Sherwood, R.T., C.C. Berg, M.R. Hoover, and K.E. Zeiders. 1983. Illusions in visual assessment of Stagonospora leaf spot of orchardgrass. Phytopathology. 73(2): 173–177.

Shoji, J., N.D. Charlton, M. Yi, C.A. Young, and K.D. Craven. 2015. Vegetative hyphal fusion and subsequent nuclear behavior in *Epichloë* grass endophytes. PloS One. 10(4): e0121875.

Smit, H.J., B.M. Tas, H.Z. Taweel, S. Tamminga, and A. Elgersma. 2005. Effects of perennial ryegrass (*Lolium perenne* L.) cultivars on herbage production,

nutritional quality and herbage intake of grazing dairy cows. Grass and Forage Science. 60(3): 297–309.

Sparks, A.H., P.D. Esker, M. Bates, W. Dall'Acqua, Z. Guo, et al. 2008. Ecology and epidemiology in r: Disease progress over time. The Plant Health Instructor. DOI: 10.1094. PHI-A-2008-0129-02.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9): 1312–1313.

Stier, J.C., B.P. Horgan, and S.A. Bonos. 2013. Turfgrass: biology, use, and management. American Society of Agronomy Madison, Wisconsin.

Stratton, S.D., and H.W. Ohm. 1989. Relationship between orchardgrass seed production in Indiana and Oregon. Crop Science. 29(4): 908–913.

Studer, B., L.B. Jensen, S. Hentrup, G. Brazauskas, R. Kölliker, et al. 2008. Genetic characterisation of seed yield and fertility traits in perennial ryegrass (*Lolium perenne* L.). Theoretical and Applied Genetics. 117(5): 781–791.

Sykes, V.R., F.L. Allen, A.C. DeSantis, A.M. Saxton, H.S. Bhandari, et al. 2016. Efficiency of spaced-plant selection in improving sward biomass and ethanol yield in switchgrass. Crop Science.

Tian, Z., R. Wang, K.V. Ambrose, B.B. Clarke, and F.C. Belanger. 2017. The *Epichloë festucae* antifungal protein has activity against the plant pathogen *Sclerotinia homoeocarpa*, the causal agent of dollar spot disease. Scientific Reports. 7(1): 5643.

Tilman, D., K.G. Cassman, P.A. Matson, R. Naylor, and S. Polasky. 2002. Agricultural sustainability and intensive production practices. Nature. 418(6898): 671–677.

Trethewey, J.A.K., and M.P. Rolston. 2009. Carbohydrate dynamics during reproductive growth and seed yield limits in perennial ryegrass. Field Crops Research. 112(2–3): 182–188.

Trethewey, J., M. Rolston, C. McGill, and J. Rowarth. 2008. Is the flag leaf important in perennial ryegrass seed production. Seed symposium: Seeds for Futures. Proceedings of a joint symposium between the Agronomy Society of New Zealand and the New Zealand Grassland Association held at Massey University, Palmerston North, New Zealand. p. 26–27.

Trupp, C.R., and I.T. Carlson. 1971. Improvement of seedling vigor of smooth bromegrass (*Bromus inermis* Leyss.) by recurrent selection for high Seed weight. Crop Science. 11(2): 225–228.

U.S. Department of Agriculture, Agricutultural Marketing Service (AMS). 2018. Certificate management system. https://apps.ams.usda.gov/CMS/CropSearch.aspx.

Van Der Plank, J.E. 1963. Plant diseases: epidemics and control. Academic press, New York.

Vandeleur, R.K., and G.S. Gill. 2004. The impact of plant breeding on the grain yield and competitive ability of wheat in Australia. Australian Journal of Agricultural Research. 55(8): 855–861.

Vogel, K.P., and J.F. Pedersen. 1993. Breeding systems for cross-pollinated perennial grasses. Plant Breeding Reviews.11: 251–274.

Waldron, B.L., N.J. Ehlke, D.L. Wyse, and D.J. Vellekson. 1998. Genetic variation and predicted gain from selection for winterhardiness and turf quality in a perennial ryegrass topcross population. Crop Science. 38(3): 817–822.

Waldron, B.L., J.G. Robins, M.D. Peel, and K.B. Jensen. 2008. Predicted efficiency of spaced-plant selection to indirectly improve tall fescue sward yield and quality. Crop Science. 48(2): 443–449.

Wäli, P.R., M. Helander, O. Nissinen, and K. Saikkonen. 2006. Susceptibility of endophyte-infected grasses to winter pathogens (snow molds). Botany. 84(7): 1043–1051.

Wanyera, R., J.K. Macharia, S.M. Kilonzo, and J.W. Kamundia. 2009. Foliar fungicides to control wheat stem rust, race TTKS (Ug99), in Kenya. Plant Disease. 93(9): 929–932.

Warringa, J.W., and A.D.H. Kreuzer. 1996. The Effect of New Tiller Growth on Carbohydrates, Nitrogen and Seed Yield per Ear in *Lolium perenne* L. Annals of Botany. 78(6): 749–757.

Warringa, J.W., and M.J. Marinissen. 1997. Sink-source and sink-sink relations during reproductive development in *Lolium perenne* L. NJAS Wageningen Journal of Life Sciences. 45(4): 505–520.

Warringa, J.W., P.C. Struik, R. De Visser, and A.D.H. Kreuzer. 1998. The pattern of flowering, seed set, seed growth and ripening along the ear of *Lolium perenne*. Functional Plant Biology. 25(2): 213–223.

Welty, R.E., and R.E. Barker. 1993a. Reaction of twenty cultivars of tall fescue to stem rust in controlled and field environments. Crop Science. 33(5): 963–967.

Welty, R.E., and R.E. Barker. 1993b. Reaction of twenty cultivars of tall fescue to stem rust in controlled and field environments. Crop Science. 33(5): 963–967.

Welty, R.E., R.E. Barker, and M.D. Azevedo. 1991. Reaction of tall fescue infected and noninfected by *Acremonium coenophialum* to *Puccinia graminis* subsp. *graminicola*. Plant Disease. 75(9): 883–886.

Welty, R.E., R.E. Barker, and M.D. Azevedo. 1993. Response of field-grown tall fescue infected by *Acremonium coenophialum* to *Puccinia graminis* subsp. *graminicola*. Plant Disease. 77(6): 574–575.

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer.

Wiewióra, B., G. Żurek, and M. Żurek. 2015. Endophyte-mediated disease resistance in wild populations of perennial ryegrass (*Lolium perenne*). Fungal Ecology. 15: 1–8.

Wilkins, P.W. 1991. Breeding perennial ryegrass for agriculture. Euphytica. 52(3): 201–214.

Wille, P.A., R.A. Aeschbacher, and T. Boller. 1999. Distribution of fungal endophyte genotypes in doubly infected host grasses. The Plant Journal. 18(4): 349–358.

Xia, C., N. Li, Y. Zhang, C. Li, X. Zhang, et al. 2018. Role of *Epichloë* endophytes in defense responses of cool-season grasses to pathogens: a review. Plant Disease.

Yamada, T., E.S. Jones, N.O.I. Cogan, A.C. Vecchies, T. Nomura, et al. 2004. QTL analysis of morphological, developmental, and winter hardiness-associated traits in perennial ryegrass. Crop Science. 44(3): 925–935.

Yu, J., S.M. Sharpe, A.W. Schumann, and N.S. Boyd. 2019. Deep learning for image-based weed detection in turfgrass. European Journal of Agronomy. 104: 78–84.

Yue, Q., C.J. Miller, J.F. White, and M.D. Richardson. 2000. Isolation and characterization of fungal inhibitors from *Epichloë festucae*. Journal of Agricultural and Food Chemistry. 48(10): 4687–4692.

Zhang, T., E.G. Lamb, B. Soto-Cerda, S. Duguid, S. Cloutier, et al. 2014. Structural equation modeling of the Canadian flax (*Linum usitatissimum* L.) core collection for multiple phenotypic traits. Canadian Journal of Plant Science. 94(8): 1325–1332.

Żurek, G., S. Prończuk, and R. Legumes. 2007. Relation between seed yield potential and turf quality in *Poa pratensis* L. 27th Eucarpia symposium on Improvement of Fodder Crops and Amenity Grasses. p. 14.